

Assessing Waldegrave's big test

Expectations are high for the British government's promised White Paper on science, now due to appear next week. Here are some criteria by which its recommendations should be judged.

ANY day now, Mr William Waldegrave, the minister in the British Cabinet responsible for science, will publish what promises to be the most significant government statement about science for 20 years. Further recommendations are unlikely to influence the forthcoming White Paper on the organization of research. But, exactly a year after Waldegrave invited suggestions, it remains clear that the outcome will be a stringent test of the government's commitment to maintain and develop the strength of British science. Thus, in deference to the stubborn enthusiasm of the education minister, John Patten, for assessing children at school, here is a scorecard by which Waldegrave's efforts should be judged.

Commitment. Look first for a clear statement that the government recognizes that Britain needs a strong scientific base. (Do not be distracted by platitudes; negative marks or grades should be given for the recitation of past glories.) Equally important, look for a demonstration that the government has the political will to put its commitment into practice. A high-level committee, perhaps chaired by the prime minister and including all ministers whose departments spend heavily on research, would go a long way to achieving both objectives. Award lower scores to schemes for coordination either by centralizing political authority on the Office of Science and Technology (OST) or by leaving it to civil servants alone.

Advice. Award significant marks also to successful restructuring of the operation of government advisory committees. It seems that the days are numbered of both the Advisory Council on Science and Technology (ACOST, whose independence from government is suspect) and the Advisory Board for the Research Councils (many of whose responsibilities could be taken on by OST). Far preferable would be a single truly independent advisory committee, with a small non-career research secretariat, and the authority to set up *ad hoc* working groups on areas of policy chosen by itself; it would run in parallel with the government's own science policy committee. Award high marks for the appointment to the committee of individuals distinguished professionally and for their independent views; negative scores for tokenism and the representation of special interests.

Strategy. British research needs a mechanism for forging agreement on strategic priorities, whether on the expectations of spending or on areas of work. Judge the proposals in the White Paper (if any) by these tests: the technique of 'research foresight' may help to focus attention on the importance of a strategy, but the important need is that the

exercise should not be excessively prescriptive. Extrapolating present trends, or more conventional techniques of social and economic forecasting should earn scores at the bottom of the range. The best ways of setting priorities are those that carry commitment (as in France over the past decade) and flexibility (so that they are not straitjackets for either academic or industrial programmes).

Careers. Something must be said (and done) about careers in British scientific research. Three problems obtrude: the lack of sufficient university posts for young scientists (likely to engender in a few years a lack of candidates qualified for top appointments); the insecurity of those employed on a succession of short-term contracts; and low pay of science graduates. Award high scores for workable solutions of any of these. A parallel career structure for university researchers should score well; the more radical proposal to pay 'bounties' to scientists at the end of short contracts to compensate them for the extra risk would deserve a bonus. Platitudinous declarations that the market will look after career prospects should be taken to signify that the British government has not understood how its predecessors over 14 years have damaged British science by exploiting the innocence of researchers in established posts in playing the labour market; the result is the flight of young people to occupations other than science.

Organization. The White Paper will have much to say about the organization of the research councils, through which most of the government's extramural spending on civil science is channelled. But only one of the five needs close attention. The Science and Engineering Research Council (SERC) has the challenge of supporting research both for wealth creation through new technology and as an academic pursuit whose tangible benefits are uncertain. A rational solution of the dilemma should earn high marks, but hiving off an Engineering Research Council would be mere tokenism. Award no score, or a negative one, for the mass privatization of the council's research institutes, as proposed by ACOST: that way lies disaster.

Basic research. Waldegrave's assessment will hang on his proposals for protecting basic research in a political environment that regards wealth creation as the scientific community's first priority. Award no marks for a mechanical formula or schemes for supporting astronomy and high-energy physics out of the pot that also supports the arts. Formalizing a split between basic and strategic research would be similarly a black mark. Better that the level of



support for basic science should be related to the need for and value of the graduates trained in research. But other forms of protection, if imaginative, may command high grades.

Departmental research. One of the most contentious issues of the past year is what the White Paper should say about research in government departments. Waldegrave will score well if he has found a way of injecting into the culture of each department an awareness of its dependence on long-term research, breaking the prevailing habit of short-termism. (Rothschild tried and failed.) But departments must not be shackled to research agendas that do not meet their own needs; again, there must be a balance between scientific effectiveness and political accountability; Waldegrave will be judged on the skill with which he proposes to manage that.

Public understanding. Like motherhood, the public understanding of science is an admirable cause and, in Britain, is in fashion. Award no marks if the White Paper proposes using public funds for artificial schemes to glamorize science. The real need can be met only in the schools. Better curricula for science students and better salaries for their teachers are not directly in Waldegrave's gift, but he needs leverage. He will deserve extra marks if his White Paper has constructive proposals for giving non-science school students a sense of what science is about.

Europe. International cooperation at the European level is prospering reasonably well, but research programmes run through the European Commission of the European Communities in Brussels still tend to rate high in administrative effort and low in value for money. If the White Paper proposes new ways of increasing the return on the British investment in European science, it will deserve high marks. Award bonus score for imaginative schemes for ventures further east.

World science. From high-energy physics to the environment, global research has become a reality. How should Britain cope? Waldegrave should say something. Arranging to protect the international science base from exchange-rate fluctuations would be a good start. The wider need is for greater rationality in planning and financing strictly international ventures. Waldegrave may (to the benefit of his scorecard) support the efforts of bodies such as the Organization for Economic Cooperation and Development to this end. Tough-minded invigilation of international programmes mounted under UN treaties would benefit even the poor countries, not to mention their sponsors.

Presentation. Overall, Waldegrave will get high marks for crisp and clear presentation; the way his message is conveyed will be as important as its content. But empty promises and unrealistic commitments will score badly. That means matching commitments by adequate funding. The British research community would welcome a significant increase in the science budget, but would prefer that its goals should be realistic. Flair is no substitute for funds. Some extra funds now, and a commitment to back declarations of intent with the costs that will arise, will be another test of the government's commitment to whatever Waldegrave proposes. He will deserve a high score if he has won the rumoured battles with the Treasury on this account. □

Basic research revisited

Although the US National Science Board knows what it wants to say, it has trouble making itself clear.

FOR 40 years, the US National Science Board has quietly tried to make sure that its charge, the US National Science Foundation (NSF), stuck to a simple but vital goal: to maintain the excellence of academic research in the United States. But last week, as the board struggled to produce a two-page public statement on the value of basic research, it became clear how much more complicated that job has become. The statement, adopted in principle but awaiting final editing, is a response to the Commission on the Future of NSF, which last autumn suggested that the board should play a larger role in setting national science policy (see *Nature* 360, 285; 1992). An internal task force headed by Frank Rhodes, president of Cornell University, has been brooding on the board's response, and has decided to begin with a new definition of NSF's mission.

That, plainly, has not been easy. In today's political climate, a nation's scientific work force is expected not merely to throw back the frontiers of ignorance, but also to revive a sluggish economy, eliminate a trade deficit and raise the standard of living, as well as to cure the sick and provide protection from military attack. Although the science board knows better than to promise that a \$3-billion agency can have much impact on an economy three orders of magnitude bigger, much less that research can make a sluggish economic engine run smoothly, its members cannot afford to ignore the devotion of President Bill Clinton and the US Congress to precisely that objective. Would NSF's budget continue to grow if they were indifferent to political reality?

The unfortunate result is that the science board has discarded perfectly understandable concepts such as basic and applied research in favour of the latest buzzwords, notably 'curiosity-driven' and 'strategic' research to distinguish between academic research and that more deliberately intended to pave the way for new technology. The terms may be familiar elsewhere, in Britain for example, but they will inevitably make the science board sound a little like Humpty Dumpty who, in *Through the Looking Glass*, declared that "when I use a word, it means exactly what I choose it to mean — neither more nor less". By recategorizing some basic research as 'strategic', which sounds better, the science board risks either alienating its scientific constituency or, possibly worse, leaving it to wonder whether NSF has lost its way.

It is ironic that the board began this exercise with precisely the opposite goal in mind: that of defining NSF's mission in a way that would make clear the importance of fundamental research and, in so doing, of protecting its budget from assault by those who would have NSF chase after whatever 'national need' is most pressing at the time. It is not too late for the board, in explaining what its document really means, to stand up for the principle that the long-term value of a robust scientific enterprise is more than sufficient justification for the continued support of NSF. □

US Army to use peer review in breast cancer programme

Washington. The US Army says it will use a system of peer review modelled on that of the US National Institutes of Health (NIH) to allocate \$210 million for breast cancer research, allaying fears that it might ignore the wishes of the biomedical community in spending its one-time appropriation.

The approach is described in a report published last week by the Institute of Medicine (IOM) recommending that the money be used for investigator-initiated grants, fellowships and proposals to strengthen the research infrastructure to battle against a disease that will this year kill 46,000 women in the United States. Members of the IOM panel say that it could also help to defuse the issue of 'pork-barrel' science by serving as a model for other US government agencies given money for programmes not directly connected to their missions.

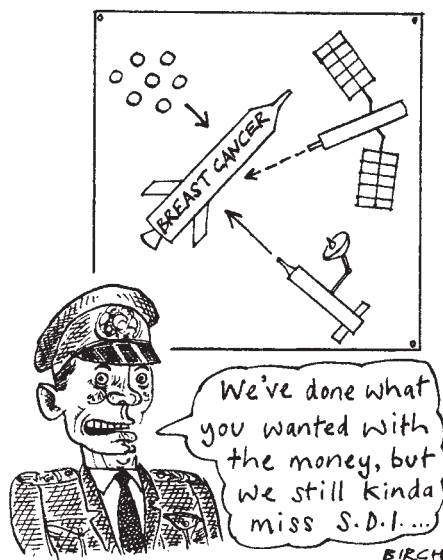
"All of it makes sense to us", says Colonel Fred Tyner, deputy to General Richard Travis, commander of the US Army Medical Research and Development Command at Ft Detrick, Maryland, about the report his office requested. Writing to the panel, Travis said that the report "achieves all of our goals", namely, to obtain advice on the types of programmes most likely to reduce morbidity and mortality in breast cancer and on ways to select the best proposals.

The Army must report to Congress on its plans by 1 June, in accordance with rules requiring it to commit the money by 30 September 1994 and prohibiting its use for new facilities or for large programme grants. Senator Tom Harkin (Democrat, Iowa), whose amendment resulted in an eightfold increase in 1993 for a small Army programme to combat breast cancer, said last week that he hopes "these recommendations will ensure that this money is well-spent".

The centrepiece of the IOM recommendation is a programme of investigator-initiated research costing at least \$150 million, most of it spent on four-year awards that would be similar to a standard NIH grant. It estimates that 160 such awards could be made, some to people already being supported by the federal government. It proposes 30 smaller, shorter awards for investigators with "innovative and exploratory" proposals — in other words, speculative work with a potentially large return — and 25 awards to young investigators with little experience.

Although the panel did not list specific types or areas of research, it said that the work should address such questions as the genetic component of the disease, changes in cellular and molecular functioning, the impact of risk factors at the molecular level

and the application of such knowledge "to improve detection, diagnosis, prevention, treatment and follow-up care". It emphasized that some of the work could involve other types of cancer or fundamental ques-



tions that shed light on the disease. It is silent on the proper balance between clinical and laboratory research, believing instead, as panellist Harold Varmus of the University of California, San Francisco, explains, that "quality should be the guiding principle".

The panel also proposes a \$27-million programme of training and recruitment to attract talented scientists of varying backgrounds into the field. "We do not feel that NCI [the National Cancer Institute] is doing a bad job", says panel chair Suzanne Oparil, a cardiologist at the University of Alabama, Birmingham, "but we want to take advantage of this new money to try something new. Kay Dickersin, an epidemiologist at the University of Maryland School of Medicine, says that "we want to bring in new people and test out some crazy ideas".

One element would offer 50 paid 'instant sabbaticals' to scientists in mid-career, and another would sponsor a dozen or more interdisciplinary meetings. There would also be money for predoctoral traineeships at universities and portable fellowships at both the pre- and post-doctoral levels. "We want everybody to realize that breast cancer is a big problem and we want to encourage them to get involved", says Varmus.

Regardless of the type of programmes offered, there is general agreement that the crucial element is the calibre of the people who will run them and who will administer the peer-review panels. Panel members say that the Army must go outside its ranks to find people with the necessary experience in the field and familiarity with the peer-review process, and Tyner agrees. "We're not breast-cancer experts", he says. "We need to find good people, and I don't expect them already to be on board." **Jeffrey Mervis**

Russian budget allows for more grants

Moscow. The government-supported Russian Foundation of Basic Research has given out ten times the expected number of grants to individual researchers after the Russian government significantly increased its 1993 science budget.

The new foundation (see *Nature* 361, 485; 1993) has awarded 3,000 grants to groups of seven to nine scientists in six broad areas; physicists received 20 per cent of the total, with the remainder divided equally among the fields of mathematics, chemistry, biology, earth sciences and the humanities. The larger number of awards was made possible by a decision of the government to increase more than threefold the initial request of 250 billion rubles (US\$500 million) for science.

Although the foundation went to some length to conduct an independent review of the proposals it received, including on-site visits to many institutions, it has been criticized for showing favouritism to those who

served on its review panels. Alexander Zakharov, one of the leaders of the professional union of scientists at the Russian Academy of Sciences, says that, as a rule, scientists involved in the selection process not only received grants themselves but also secured funding for coworkers. The review panels were headed by members of the foundation's governing council, who were ineligible for the awards.

The president of the foundation, Andrei Gonchar, does not dispute such allegations and sees nothing wrong with the practice. "I believe that an invitation to a scientist to become an expert means *a priori* that such a scientist is worthy of support if he submits a grant proposal", Gonchar says. The same is true for officials at the Russian Academy of Sciences, including president Yuri Osipov, who Gonchar says were chosen not because of their prominent positions but because they are good scientists.

Vladimir Pokrovsky

UK national laboratories face new threat of privatization

London. Three years after being turned into 'executive agencies' and told to pay their way entirely through contracts with government departments and private industry, the five national laboratories run by Britain's Department of Trade and Industry (DTI) are once again under scrutiny.

Michael Heseltine, the president of the board of trade, has said that the laboratories must undergo a 'cultural change', with the

of the laboratory's 250 scientists will lose their jobs and the rest will be moved to the DTI-owned company AEA Technology at Harwell, itself scheduled for privatization.

The review is also likely to make little difference to the NEL in Glasgow, which provides testing facilities and other services for the engineering industry. The laboratory is scheduled for privatization in 1995.

Officials at NEL say that a variety of new administrative procedures introduced since 1988, when the government tried unsuccessfully to privatize it, have already created the shift towards the industrial values being demanded by Heseltine. One important change is demanding greater accountability from individual projects.

But the privatization of NEL is still strongly opposed by the Institution of Professionals, Managers and Specialists, the trade union representing staff members. Union officials claim that the previous attempts to streamline the organization during its ef-

forts to find a buyer resulted in the loss of 400 jobs, and that this number could grow.

Perhaps the biggest question mark hangs over the future of the National Physical Laboratory at Teddington, Britain's equivalent of the US National Institute of Standards and Technology. NPL became an executive agency in 1990 and since then has been required both to find an explicit customer for each of its programmes and to balance its books without additional governmental support. But DTI, which last year provided contracts worth about £40 million (\$62 million), remains its main customer.

Budget cuts last year forced DTI to reduce some of NPL's activities by 20–50 per cent, in particular research on measurements in radiation and acoustics. Further cuts are expected; NPL director Peter Clapham has confirmed that overall funding from the DTI is expected to fall by as much as 30 per cent over the next three years. The reductions could lead to the loss of 100 of the NPL's 850 staff.

Industry is becoming increasingly concerned. Last week, the president of the Institution of Physics, Clive Foxell, wrote to the trade minister, Edward Leigh, saying that

the institute's industrial affiliates are worried that the cuts could jeopardize NPL's usefulness and its international credibility. According to Foxell, although some companies have criticized NPL for setting its fees too low and for at times spending too much time on projects that are no longer productive, they also feel that "NPL is carrying out important state-of-the-art work in fields such as optoelectronics which is vital to the success of companies working in the field".

Others claim that senior managers in industry and government have yet to realize the significance of the cuts because much of NPL's work is better known to technical staff. There is also concern that full-scale privatization could be counterproductive. "There is a danger that other countries, which provide strong government support to their measurement laboratories, will treat us as a joke," says Bill South, a non-executive director of Philips Electronics and a member of NPL's steering committee.

DTI officials are expected to respond to the criticism with a survey of industrialists. Laboratory officials and scientists hope that the results will influence the review being commissioned by Heseltine and prevent any hasty action that could inflict irreparable damage.

David Dickson

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Acoustics research is one of many areas that NPL director Peter Clapham, right, has been forced to cut.

possibility of full-scale privatization, to bring themselves in line with the values and management practices of industry. With this in mind, he has asked outside consultants to study the future of five DTI laboratories: the National Physical Laboratory (NPL), the Laboratory of the Government Chemist and the National Weights and Measures Laboratory, all at Teddington outside London; the National Engineering Laboratory (NEL) in Glasgow and the Warren Spring Laboratory near Stevenage. In response, research staff complain that budget cuts and staff reductions have hindered their ability to provide long-term support to British industry.

The review appears to have come too late for Warren Spring, the government's main centre for environmental technology, which conducts work ranging from effluent measurement to industrial biosafety. Last week, Heseltine is reported to have told the director of the laboratory, Doug Cormack, that it would be closed down.

DTI says officially that it has made no firm decision about the future of the Warren Spring, and staff at the laboratory are still hoping that Heseltine can be persuaded to change his mind. If the decision stands, 150

US companies settle EPO patent suit

Washington. Two leading US biotechnology companies, Amgen Inc. and Genetics Institute Inc., settled an old score last week with an agreement on damages in a six-year patent dispute involving erythropoietin (EPO), one of biotechnology's most lucrative drugs. It is used to treat anaemia by stimulating the body's own production of red blood cells. The settlement resolves all patent disputes over EPO in the United States between the two companies.

Under the settlement, Genetics Institute will pay Amgen US\$14 million in damages. In 1991, Genetics Institute took an \$11-million charge against earnings after a federal appeals court upheld Amgen's EPO patent and ruled that Genetics Institute's patent was invalid (see *Nature* 350, 99; 1991). Licensee Boehringer Mannheim GmbH has also agreed to reimburse Genetics Institute for a portion of the costs. Sales of Amgen's EPO product, approved in the United States in 1989, reached \$507 million in 1992.

In a separate patent settlement, Baxter International, Genetic Institute's worldwide licensee for its recombinant Factor VIII product (Recombinate AHF), has agreed to pay \$105 million to Rhône-Poulenc Rorer and the Scripps Research Institute of La Jolla, California. Genetics Institute will share the cost of future royalties paid to Rhône-Poulenc Rorer on Baxter's sales of Recombinate AHF, a drug used to treat haemophilia A.

Diane Gershon

Market for superconductors is expected to soar by 2020

Tokyo. Industrialists in Japan, Europe and the United States involved in superconductor research believe that, given sufficient government support for research and development, the market for superconductors could rise from its present level of US\$1.5 billion to \$150–\$200 billion by 2020. The projections, made at the second International Superconductivity Industry Summit (ISIS) held in Hakone by Mount Fuji in

consensus on both the projected size of the total market and the relative share of individual applications.

The present market of \$1.5 billion consists entirely of products made from conventional low-temperature superconductors, in particular magnetic resonance imaging devices for medical use. But by 2020, products made with high-temperature superconductors will comprise the dominant share, predicts CSAC representative Alan Lauder of DuPont.

ISIS members admit that there are uncertainties in their estimates. "Superconductivity may well develop a major business through an unexpected technological breakthrough in an unrelated field", according to ISIS, in the same way as the market for semiconductor lasers grew rapidly because of compact-disc players. And the discovery of superconductors with even higher critical temperatures (T_c) could result in even larger markets, says Lauder.

But much of this growth will depend on government policies. Achieving the projected market size will require an investment of "hundreds of millions of dollars over the next 30 years", says Lauder, some of which must come from governments.

In the United States, that investment has grown from \$40 million in 1986 to \$250 million in 1992, much of it for low-temperature superconducting magnets for the Superconducting Super Collider (SSC). Similarly, Japan increased its spending dramatically after the discovery of high-temperature superconductors in 1986 to the present level of about ¥18 billion (US\$162 million). Investment in Europe is on the same order of magnitude as in Japan, says Gordon Donaldson of the University of Strathclyde in Scotland.

But the future does not look so bright. Britain's Department of Trade and Industry has just terminated a programme that provided about £8 million (US\$12 million) a year, and Japan's Science and Technology Agency (STA) will reduce its support for an important project in the field from ¥2.1 billion this year to no more than ¥1.3 billion next year, although STA officials hope to start a second phase in 1995. ISIS members in the United States admit that one of the purposes of the market projection is to persuade the US government to continue its investment in superconductor research regardless of the fate of the SSC.

Next year, at a meeting in Europe, ISIS plans to arrange practical demonstrations of superconductor devices. The first projects are likely to be national ones, says Piel of Germany, followed by international collaboration.

David Swinbanks

China struggles with home-grown 'Gordon conference'

Beijing. Five years ago, the Chinese State Science and Technology Commission started an annual seminar modelled on the Gordon Research Conferences held every summer in New Hampshire in the United States. The seminars, called the Xiangshan ('Fragrant hills') Science Conference after a picturesque site in the foothills west of Beijing, were intended to foster the free exchange of ideas among different disciplines.

Last month, as prominent scientists such as Zhou Guang-Zhao, president of the Chinese Academy of Sciences, gathered at Xiangshan, it was clear that the concept has failed to take root. The traditions of science in China militate against such a free-flowing debate among those in different fields and of different rank, and the result, according to participants, has been discussions that can charitably be described as dull and non-substantive.

The rules are not to blame. Participants are told to speak their minds and government officials, in a remarkable departure from common practice, take part only as individual scientists. Those who have attended say that the procedures are quite helpful.

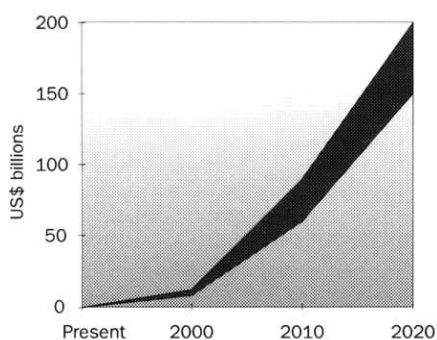
But Chinese scientists are not accustomed to open debate. Public disagreements are regarded as rude in a culture that reveres seniority. In addition, without an effective system to protect intellectual property, Chinese scientists are reluctant to talk about new findings for fear that others will steal their ideas. Cases of alleged infringement fill the courts, and even groups working on similar topics within the same institution refuse to share data. The Chinese refer to such communications problems as "internal friction", but efforts by government officials to bring about change have had little effect.

Part of the problem rests with training. Educated in a university system drawn from the rigid Soviet model, most researchers have a very narrow view of science and a limited knowledge of fields outside their speciality. As a result, they can contribute little to a discussion that ranges across several disciplines. Instead, participants choose topics that do not require specialized knowledge, for example 'science in the next century', and in which they are unlikely to be proved wrong.

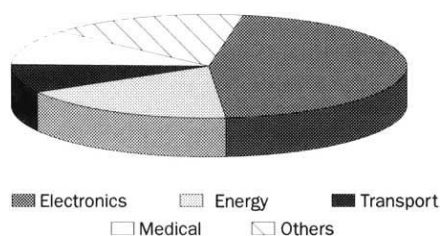
Despite questions about the value of the conferences, the state commission seems determined to support them. Zhou Guang-Zhao would like to see them continue, and the commission has already allocated money for a sixth meeting next year. Given the dissatisfaction among scientists, however, organizers may have to face the fact that good intentions alone are not enough to sustain this annual gathering.

You Qin Li

The promise of superconductors...



... and their uses



Japan, are intended to persuade national governments to continue their support for superconductor research beyond the several hundred million dollars a year being spent in the three regions.

ISIS, made up of industrialists and researchers from Japan's International Superconductivity Technology Center (ISTEC), the US Council on Superconductivity for American Competitiveness (CSAC) and the recently formed Consortium of European Companies determined to use Superconductivity (CONECTUS), made the projections on the basis of information from companies in each region. Projections were converted to a worldwide figure on the basis of gross national product. Helmut Piel of the Bergische Universität Gesamthochschule Wuppertal in Germany, who with Sir Martin Wood of Britain's Oxford Instruments helped to create CONECTUS in March, says that he was pleasantly surprised at the

Berlin universities jockey for position after reunification

Munich. German reunification has created fierce competition for money, students and courses between three large universities in Berlin. The outcome of that rivalry will shape the health of science in Germany's once and future capital city.

The Free University and the Technical University in the west part of Berlin share the same pot of funding with east Berlin's Humboldt University, although a generation of isolation spawned different teaching and research structures and the duplication of many courses.

After reunification, the German science council, the Wissenschaftsrat, evaluated Humboldt with the idea of bringing eastern institutions into line with their western counterparts. But long after most east German universities completed the process, the Humboldt continues to fight against many decisions that it feels are unfair. At the same time, the two western universities see the 'renewal' of the Humboldt as a drain on their own resources.

The Humboldt, Berlin's oldest university, was established in 1810 by Wilhelm von Humboldt, whose theory of the unity of teaching and research remains in the German constitution, and soon became one of Europe's leading universities. Then called the Friedrich Wilhelm University, it was suppressed by the Nazis and finally closed in 1944 after 70 per cent of its buildings were destroyed. In 1945 it was reopened in the Russian sector of occupied Berlin and in 1949 it was given its present name. Today it has nearly 20,000 students.

The King's Technical College was founded on the site of the 'Bergakademie' (established in 1770) and achieved university status around the turn of the century. In the 1930s it became strongly associated with national socialism, and all Jewish academics and Nazi dissidents were expelled. After the war it was reopened as the Technical University to dissociate itself from its past. Its current enrolment is 37,000.

Politics also led to the creation of West Berlin's second university. In 1948, at the beginning of the Cold War, the Humboldt expelled western students, who agitated for their own university. The result, financed by the military government in the US sector of Berlin, was the Free University, which now has 61,000 students.

The communist government separated research from higher education and placed it in institutes belonging to the national Academy of Sciences. The 1990 Higher Education Renewal Programme required research to be reintegrated into eastern university systems (see *Nature* 362, 775; 1993), leading to the construction of new facilities on

top of the extensive renovation required to raise the infrastructure to western levels.

The cost of this restructuring has taken its toll on Berlin's budget for higher education. The Free University has lost 55 faculty posts and an additional 110 posts have been temporarily frozen. The Technical University has lost 40 posts and 80 have been frozen. The already high ratio of students to staff at these universities (66:1 and 70:1 respectively) is likely to increase because, by law, enrolment in some subjects can be reduced only when posts are eliminated, not simply frozen. The western universities also attract students from the eastern sector who prefer to study in the better facilities that the west offers.

The need to coordinate courses is another problem. In early 1991, attempts were made to close down departments at the Humboldt on the grounds that it was being restructured anyway. But such back-door decisions by the Berlin senate, the legislative body for the local government, were successfully challenged in the courts.

In some departments, such as law, history and philosophy, the process of closure was already under way and posts have been reassigned. As a result of the court ruling, universities in some cases are paying for two people to occupy the same faculty post.

Academic departments may be closed only after an evaluation by the Wissenschaftsrat of equivalent departments in all three universities. But the Humboldt still feels at a disadvantage because a history of inadequate financing has made it difficult to compete with western departments.

In several cases, duplication has been eliminated by combining departments and housing them at one site. But the unequal wages and different scientific qualifications have caused social problems and generated complaints from those at the Humboldt. For example, the Free University was chosen as the site for a combined veterinary medicine department despite the fact that the Humboldt's department was rated higher. The Berlin senate decided that its rural location was more appropriate for handling large animals and that its facilities were better. And last summer the Berlin senate combined two electronic engineering departments on the understanding that the Humboldt could continue to send students through the new department. But its department has effectively been closed after the senate failed to keep its promise.

Other unions have been less acrimonious. For example, in October agricultural sciences were transferred from the Technical University to the Humboldt, with food technology moving in the other direction.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Differing pasts merge into a common future for, from top, the Free, Technical and Humboldt universities.

But attempts to rationalize Berlin's three costly university clinics, the Humboldt's Charité with 2,450 beds and the Free University's two clinics, the Rudolf Virchow with 1,700 beds and the Steglitz, with 1,350 beds, were attacked so strongly by each university that in March the senate decided to make the two larger clinics the same size as the smallest one. Along with that reduction, the number of students will decline from more than 1,000 to 600 and the number of lecturers from 400 to 330, with the senate dictating the areas of specialization.

Although a limited level of scientific interchanges between east and west existed despite the Wall, the new opportunities to broaden that cooperation are not being fully exploited. Instead, university researchers prefer to form links with research institutes because the competition for jobs and funding has proved to be a disincentive for cooperation.

Cornelia Koob

Deep divisions complicate global environment meeting

Washington. Representatives of 60 governments will gather in Beijing this weekend to discuss their disagreements about the structure and funding of an international agency set up to help poor countries tackle global environmental issues.

The Global Environment Facility (GEF) — an offshoot of the United Nations and the

During the pilot, the European countries promised \$800 million and the United States, Japan, Canada and Australia \$400 million for projects to be selected by each country. But none of the \$150 million pledged by the United States, for example, has been allocated, and European officials say that future participants will be required to make a more

definite commitment. GEF is expected to need \$3–\$4 billion over the next five years.

Although US President Bill Clinton is said to be keen to reverse the policies of his predecessor and to contribute to GEF, administration officials are refusing to do so until GEF meets a list of conditions that would open up the process. Congress has decreed that the \$30 million appropriated for GEF this year will

not be spent unless those changes are made by September, and Lawrence Summers, under-secretary of state for international facilities, told a congressional committee two weeks ago that the concerns so far

had not been dealt with.

The chief criticism of GEF is that it is inheriting not only the secrecy and bureaucratic slowness of its administrative parent, the World Bank, but that it is also diverting the World Bank from environmental issues in its other programmes.

Environmental groups want the US government to demand more openness at the GEF before spending any money. But they say that it will be difficult for the United States to champion reform at Beijing when it has not contributed to GEF. They also say that the United States must explain exactly what it wants GEF to accomplish.

"We need a clearer picture of exactly what the US agenda is for reforming the Global Environment Facility", says Glenn Prickett of the Natural Resources Defense Council in Washington. "The US position doesn't exist yet."

European unhappiness with US policy is likely to surface at Beijing. "The United States has been incredibly active on the minutiae of GEF", says one European staffer at GEF, "but so far they haven't put in a bean." The attitude of the US delegation will determine if the mood will improve, with Prickett hoping for "a new tone" and a more productive relationship with the rest of the world.

The meeting will also examine ways in which non-governmental organizations (NGOs) in both donor and recipient countries can become involved in GEF's operations. An NGO meeting immediately before the main event will give these bodies a chance to be heard, and the conference will receive a report on how they could be more deeply involved.

Colin Macilwain

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Reforestation Africa is one goal of GEF.

World Bank — is nearing the end of a three-year pilot phase during which it has pledged more than \$700 million to support 113 projects, ranging from biodiversity conservation in Cameroon to the management of waste from ships in the Caribbean. At last year's Earth Summit in Brazil, it was agreed that the GEF — subject to certain reforms — should also serve as the main vehicle to finance projects stemming from the treaties on biodiversity and climate change.

That role is, however, threatened by divisions between the United States and European countries on how to pay for GEF and between the rich and poor countries on how it should make its decisions. The prospects for resolving those differences at Beijing are not good, officials say, although some sort of agreement is expected before the pilot phase runs out in December.

At the top of the agenda at Beijing will be reform of the decision-making structure of GEF. The poor countries are seeking a system that gives one vote for each participant, the rich a system weighted according to a country's financial contribution. The conference will discuss a structure that will give each side an effective veto by requiring a 'double majority' — of both the number of countries and the amount of money contributed — but no decision is expected until after meetings in Washington in September and Geneva in December.

The other major item of discussion will be financing GEF after its pilot phase ends.

Australia rethinks technology transfer

Sydney. Australia's largest research organization, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), is dismantling the commercialization company it set up in the early 1980s.

The company, known as Sirotech, will be reduced to a core of legal and patent specialists; project officers who supervised specific projects will be reassigned to the organization's far-flung laboratories.

The reorganization is not due to any particular dissatisfaction with Sirotech, formed as part of an effort to get closer to the industrial and agricultural sectors the CSIRO was set up to serve. Although Sirotech has not missed any major opportunities, it cannot claim any large-scale successes.

The company has assisted in a number of projects to commercialize the results of research at CSIRO, which has a total annual budget of A\$670 million (US\$478 million) and is Australia's largest employer of scientists. Those major projects include the Sirosmelt and related Isasmelt metal smelting process now being used to smelt copper at Mt Isa mines in the northern state of

Queensland and the exploitation of new simulation technology for the handling of air traffic.

However, CSIRO found that having a central company for an organization with 33 divisions spread all round Australia was far from ideal. A CSIRO spokesman said that companies interested in marketing a product preferred to deal directly with the scientist concerned rather than through a Sirotech project officer based elsewhere. CSIRO receives 30 per cent of its funding from non-government sources.

The hope is that moving the project officers physically closer to the scientists will bring both under the direct supervision of the organization's line managers, who are scientists. John Stocker, CSIRO's chief executive, said that is a goal of the organization's board.

The company has a staff of 45 and a budget of A\$4 million a year. The details of who will move to the operational sites and how many will remain will be decided in negotiations now under way.

Mark Lawson

US bill supports research in marine biotechnology

Washington. US university marine biologists may soon have a major new source of funds under a proposal moving rapidly through Congress to create a marine biotechnology research programme. Although legislators may not be willing to provide the \$20 million that the bill authorizes for the 1994 fiscal year, which begins on 1 October, proponents say that the measure would strengthen the country's marginal efforts to mine the seas scientifically for commercial gain and would send a "wake-up call" to the National Oceanic and Atmospheric Administration (NOAA) to encourage research in this area.

Last week, a subcommittee of the Merchant Marine and Fisheries Committee in the US House of Representatives approved the Marine Biotechnology Investment Act (HR 1916) introduced by the committee chairman, Representative Gary Studds (Democrat, Massachusetts). A vote is likely as early as next week by the full committee and by the House some time next month, by which time the Senate may have scheduled hearings on a companion bill expected to be introduced by Senator Ernest Hollings (Democrat, South Carolina). Hollings, a long-time supporter of the field, is influen-

tial as the chairman of both the Commerce, Science and Transportation Committee, which has jurisdiction over NOAA, and the relevant appropriations subcommittee that controls the agency's budget.

The new programme would be administered by the National Sea Grant College Program, which distributes \$40 million to 29 university-based offices around the country. A national review panel would judge the applications and make its decisions on the basis of scientific quality rather than geography. In the past, Sea Grant has been criticized for the uneven quality of its awards, a problem exacerbated by a budget that has remained essentially flat for a decade.

Sea Grant officials say that their programme, was one of the first to recognize the potential of marine biotechnology and that its mission — to commercialize discoveries by linking academic researchers with companies — parallels the goals of the legislation. Projects funded by Sea Grant must obtain a third of their budget from nonfederal sources, a requirement intended to ensure that their efforts are relevant to local needs.

The biotechnology programme and national review panel would not begin work until the additional funds — \$20 million has

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Industry is looking for anti-inflammatory drugs among soft corals such as these.

been requested in 1994 and 1995 and \$25 million in 1996 and 1997 — are appropriated to avoid a dilution of existing programmes. Applications for research, education, technology transfer and related projects would be screened by the appropriate state Sea Grant offices, and those from states without such programmes would be encouraged to form partnerships.

The amount of money to be authorized is seen as giving belated recognition to the importance of marine biotechnology within the biotechnology industry. The field receives only about 1 per cent of the \$4 billion that the federal government will spend this year on biotechnology, with most of the money going to health-related projects and much of the remainder to agricultural programmes. Sea Grant spends less than 10 per cent of its money on biotechnology.

Jeffrey Mervis

French science survives cuts by new government

Paris. By losing less than other programmes, French science comes out ahead in the new government's revised budget for 1993. But prospects for increases over the next three years are bleak, and the newly appointed minister of higher education and research, François Fillon, last week confirmed that his policies, mirroring those of Prime Minister Edouard Balladur, are likely to embrace continuity rather than change.



François Fillon

Unlike his predecessor Hubert Curien, a former researcher and director of the French space agency, Fillon, who is 39, has no scientific background. An experienced politician and a specialist in military strategy, Fillon holds anti-Maastricht views (see *Nature* 363, 7; 1993) that prevented his becoming minister of defence.

Fillon criticizes the Curien administration for being preoccupied with management while neglecting to "define priorities".

But his favourites in a FF52.6-billion budget (US\$10 billion) — nuclear power, biotechnology, the human genome, space and AIDS, "areas where France has a clear lead", he says — appear little different from those of his predecessors.

One change is a decision to stop writing blank cheques for the space programme (see *Nature* 359, 659; 1992). "France needs to find a space policy, which it doesn't have any more," says Fillon, "and since France drives Europe, Europe doesn't have one either." Fillon criticizes the European Space Agency for "pouring money into manned flights, with no idea of where it is going" and says that a proposal to collaborate with Russia is a "means but not an objective".

The merger between the ministries of higher education and research removes an important obstacle to the creation of autonomous research-based universities. Balladur overturned an earlier decision (see *Nature* 362, 96, 1993) ruling out such a move because, says Fillon, "the possible synergies" outweighed the risk that the attention needed to be devoted to universities might eclipse a broader debate on research policy.

Fillon says that universities will eventually get their own budgets and exercise

control over staff but that such devolution must occur gradually. He also finds it "unacceptable that researchers do not pass on their knowledge" but prefers to offer the carrot of better prospects for promotion and pay rather than the stick of compulsory teaching duties for full-time researchers.

The new government has presented no new plans for wealth creation, but Fillon says that one possibility would be modelled on BioAvenir, a FF1,600-million fundamental research programme in health, chemicals and agriculture started in 1991 by the government and the multinational company Rhône-Poulenc (see *Nature* 353, 487, 1991). BioAvenir differs from conventional wealth creation programmes in letting the company decide what is funded.

Whatever the long-term impact of the conservative victory on researchers, they are already feeling the effects of a budget cut of 1.2 per cent imposed last week, which translates into a reduction of 4–7 per cent in laboratory spending. However, science fared well given the massive public spending cuts ordered by the new government to reverse France's worse economic slump since the Second World War.

Declan Butler

US military to sponsor lunar science mission

Washington. The US military is preparing to launch a spacecraft to the Moon for the first time since the early days of space exploration, with science along for the ride.

Although the Deep Space Program Science Experiment, nicknamed Clementine, is primarily a test of satellite technology developed for the Pentagon's Strategic Defense Initiative (SDI) programme, scientists involved with the \$80-million project

Shoemaker of the US Geological Survey in Flagstaff, Arizona, who heads a 13-member Clementine science team chosen by NASA last month, says the mission will produce a global map of lunar rock types — information that at present exists for only a few scattered sites.

Nearly four months after leaving the Moon, Clementine will fly past the asteroid 1620 Geographos using a new onboard navigation system that automatically adjusts its course in relation to the asteroid. SDI scientists will have the chance to practise tracking a target, while NASA scientists collect data on the asteroid's shape and surface.

NASA

Although the original mission was not a scientific one, Shoemaker and a handful of others helped to select the lunar orbits and the wavelengths for multi-spectral observation. Shoemaker says that he was chosen both for his experience with lunar and asteroid missions and because he held a security clearance.

It has been 35 years since the US Air Force launched one of the first attempts to reach the Moon, Pioneer 1. Since then, US Solar System exploration has been exclusively the province of

NASA, and some scientists are uncomfortable with the idea of teaming up with the military.

Clark Chapman of the Planetary Science Institute in Tucson, Arizona, who has served on several NASA advisory committees for Solar System exploration, would have preferred that the mission be led by NASA and that it should "follow some nationally established scientific goal". The SDI's planning for Clementine has been "very *ad hoc*", he says, including a hasty selection by NASA of proposed Clementine experiments. Others complain that Clementine gives the Pentagon's missile defence programme a patina of science and that it steals thunder from NASA's own plans to launch a more capable lunar orbiter — plans that have languished for a lack of funding.

Shoemaker admits that Clementine was not designed by scientists and that the spacecraft's unproven technology and the ambitious flight plan increase the risk of failure. But he sees it as "a gift" for the planetary science community that may provide valuable data. Even Chapman believes that, in the absence of a NASA lunar exploration programme, "it's better to do something than nothing".

Tony Reichhardt

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Clementine, with Earth in background, hopes to map lunar surface next winter.

expect to collect important data at little expense, including the first close-up pictures of an Earth-crossing asteroid. The National Aeronautics and Space Administration (NASA) is providing scientific advice for the mission, to be launched in January 1994.

SDI had planned to orbit a satellite around the Earth to test technology developed for anti-missile satellites, including a light-weight star tracker, battery, sensors and solar panels. But the scope was broadened after former NASA administrator Richard Truly said that his agency should collaborate on a joint programme involving new SDI technology. SDI project managers say that the extra \$30 million being spent to send the spacecraft to the Moon and to an asteroid matches the cost of launching a second spacecraft to be used as a target for testing Clementine's ability to sense an approaching missile.

The spacecraft will enter lunar orbit in late February after flying a series of orbits through different radiation environments surrounding the Earth to meet another SDI objective. For two months it will map the Moon with ultraviolet, visible and infrared sensors from an eccentric orbit that comes within 400 km of the lunar surface. Eugene

NEWS IN BRIEF

Tokyo. The endless debate about the current ban on commercial whaling is likely to drag on for at least another year. The International Whaling Commission at its annual meeting last week (see *Nature* **363**, 9; 1993) agreed to postpone any action on a revised management procedure that would allow the resumption of commercial whaling in Antarctica and instead voted to establish a working group to report back next year on a French proposal to designate the southern ocean as a whale sanctuary. **D.S.**

Munich. Germany's new research minister, Paul Krüger, intends to start negotiating immediately for more research money in next year's budget. Reserved and cautious, Krüger has declined to announce his priorities until he analyses the political situation, but he has promised to maintain the policies of his predecessors. He is expected to be particularly supportive of additional funding for east Germany which, he says, "provides only 2.5 per cent of Germany's innovative exports, despite the available potential". **A.A.**

New Delhi. The Indian government is investigating a fire earlier this month that destroyed three floors of the country's leading medical research institute. The fire at the all-India Institute of Medical Sciences in New Delhi wiped out years of results on two dozen major research projects, as well as ruining equipment and killing experimental animals. One biochemistry laboratory supplied reagents, antibodies and cell lines to researchers throughout India. Although the institute expects to receive funds to rebuild the facility and replace lost equipment, researchers blame management for having failed to install such basic precautions as alarms, smoke detectors and sprinkler systems. **K.S.J.**

London. The Japanese pharmaceutical company Eisai has promised to invest more than £60 million (US\$90 million) over the next 15 years in a new laboratory opened last week at University College, London (UCL). The laboratory will focus on the molecular and cell-biology aspects of the neurosciences, in particular diseases such as stroke, Parkinson's, multiple sclerosis and Alzheimer's. Derek Roberts, the provost of UCL, points out that the college's links with Japan stretched back to 1863, when four Japanese undergraduates — one of them a future prime minister — came to London to study analytical chemistry. The two-year-old agreement has already resulted in £14 million being spent on a new research building that will eventually become part of the college. **D.D.**

Differential salary scales

SIR — Peter Fitzgerald (*Nature* **346**, 388; 1993) draws attention to the differential salary scales for clinical and non-clinical academic staff in New Zealand universities. His letter touches on a number of tangential points, some of which demand clarification.

The clinical lecturer grade, in New Zealand as in most universities in the United Kingdom, is a non-tenured training grade equivalent to the hospital grade of registrar. The lowest tenured clinical academic grade is senior lecturer; appointment of clinical academics at this level provides approximate parity with the hospital grade of consultant, and does not imply a 10-year seniority differential over non-clinical staff.

Grant-funded staff of appropriate seniority in our university are designated professorial research fellows. We value their scholarship no less than we do those of similar seniority and salary on the university staff establishment, who are professors. The distinction is a distinction of funding source and of tenure, which for research fellows is necessarily limited to the tenure offered by the granting body. While it is not within my power to enforce the use of the honorific 'Professor' for professorial research fellows, I and many others within the university are consistent in our adoption and advocacy of it.

D. Ross Boswell

Department of Pathology,
Christchurch School of Medicine
of the University of Otago,
Christchurch, New Zealand

SIR — I am aware of two letters (John P. Gibson *Nature* **346**, 213; 1990 and Peter H. Fitzgerald *Nature* **362**, 388; 1993) discussing pay differentials at New Zealand universities for the same position occupied by a researcher with either medical or academic training. Both argue lucidly that if a position could reasonably be filled by a researcher of either background, then wage and status differentials are *a priori* both insulting and discriminatory. What surprises me is not the ill-feeling towards institutions that employ such hiring policies but that the discussion has been limited to New Zealand, when it is alive and well in the National Institutes of Health (NIH) intramural programme in the United States.

It is common knowledge within the programme that all are paid according to different schedules. These schedules and other attendant benefits can differ significantly and invariably favour the medical degree. It is also my impression that senior positions (laboratory chief and higher) are held by MD researchers in

disproportionate numbers. Although this may reflect NIH's historical placement in the Public Health Service, the existing ratio of MDs to PhDs in higher positions suggests an advancement bias.

I do not oppose paying wages sufficiently high to recruit and retain researchers with medical degrees. But for positions that do not require duties reserved by law for medical personnel, and are therefore open to best qualified candidate regardless of degree type (PhD or MD), wages, benefits and status should be equal. Other disadvantaged groups within the NIH have effectively and justifiably argued that wage and advancement discrimination reflects institutional bias and not job performance. Like their salaries and opportunities, the salaries of academics should be increased to the level that makes working for NIH attractive to medical doctors.

It is important for both medical and academic doctors to have access to research opportunities. Both training programmes provide unique qualifications to those who complete them. However, there is nothing in medical training that is inherently superior to academic training for preparing one to do research. Indeed, by the nature of their training, those with academic doctorates always have at least three more years of research experience than their medical peers.

Jack A. Heinemann

NIH, NIAID, LMSF,
Rocky Mountain Laboratories,
Hamilton, Montana 59840, USA

Stifling the media

SIR — Several writers have proposed the unproven theory that AIDS developed from polio vaccines used in Africa in the 1950s that were contaminated by simian immunodeficiency viruses from the monkey kidneys on which they were cultured.¹⁻⁴ Hilary Koprowski has threatened or launched defamation actions against some of the media outlets that have raised this theory, notably *Rolling Stone* and journalist Tom Curtis⁵.

Whatever one may think of this particular theory, the use of the courts against writers and publishers discussing scientific issues is an unwelcome development. It is likely to have an inhibiting effect on open scientific discussion⁶. This can be considered to be an analogue, in the scientific arena, of what have been called "strategic lawsuits against public participation" or SLAPPs, in which legal actions are used to harass citizens who speak out in a way threaten-

ing to developers, government bodies and other vested interests⁷.

One can imagine the effect on science if it had been considered appropriate to take legal action against Darwin for his writings on the origin of species, against writers commenting on nuclear weapons policy-making, against publishers dealing with the issues surrounding genetic engineering and so on. In such cases, it is almost inevitable that someone's views will be explicitly or implicitly brought into "disrepute". Fortunately, it is generally recognized that scientists sometimes make mistakes, are wrong, or undertake research or applications with inadvertent adverse consequences. Without learning from mistakes, they are bound to be repeated. It would be unfortunate if discussion of possible inadvertent consequences of scientific activity could be inhibited by legal action.

Brian Martin

Department of Science
and Technology Studies,
University of Wollongong,
New South Wales 2522, Australia

1. Lecatsas, G. & Alexander, J. J. S. *Af. med. J.* **76**, 451 (1989).
2. Pascal, L. Science and Technology Analysis Research Programme Working Paper 9 (University of Wollongong, NSW 2522, Australia, 1991).
3. Curtis, T. *Rolling Stone* **626**, 54-61, 106-108 (19 March 1992).
4. Elwood, B. & Stricker, R. B. *Research in Virology* (1993, in press).
5. *Science* **259**, 180 (1993).
6. Martin, B. *Philosophy and Social Action* **12**, 5-23 (Jan-March 1986).
7. Canan, P. & Pring, G. W. *Social Problems* **35**, 506-519 (1988).

Beating fraud

SIR — Much has been written about scientific misconduct allegations and the need for science to police itself better. However, it seems to me that scientists and journalists need a fundamental change of attitude towards misconduct to address this issue properly. To start with, we should refrain from the use of the inappropriate term 'whistleblower' to refer to the person who exposes potential misconduct. This unfortunate usage symbolizes the negative effect that bringing misconduct to light can cause, and careers are often irreparably damaged as a consequence. In other words, only when society does not devalue the person with enough courage to expose potential cases of misconduct will science really be able to police itself adequately and convince others that it can, which is of particular importance in today's world of accountability and limited research budgets.

Robert D. Nicholls

University of Florida Brain Institute,
Box 100244,
JHM Health Science Center,
Gainesville, Florida 32610-0244, USA

Time for scientists to pay their dues

D. A. Rees

On the eve of the announcement by the British government of its plan to reorganize the country's scientific enterprise, how should scientists see themselves in relation to the world in which they live?

WHAT scientist today is not proud of the great progress in his or her subject since the beginning of the scientific revolution and the benefits which have flowed for mankind? And yet it would seem that it is not the benefits of science which come first to the mind of lay people when thinking of our work. Fear and apprehension crowd out appreciation, let alone enthusiasm; the public no longer seems to trust us to follow the disinterested pursuit of scientific truth because it doubts our sense of responsibility in applying or misapplying the findings.

This may seem unfair when, for example, it was the military under instruction from politicians — not us — who dropped the atom bomb, industrialists — not us — who built the satanic mills, bureaucrats — not us — who were responsible for Chernobyl, gun runners — not us — who make civil wars so much more bloody, and commercial adventurers — not us — who are destroying the rain forests. The contrast between how we see ourselves and how others see us is — to borrow Arthur Koestler's metaphor — as if we scientists, sleep-walkers dreaming of splendours and benefactions of science, are now suddenly woken into a world in which a dark side of science predominates to threaten disaster and the end of all things.

In fact, we are neither the innocents of our dreams, nor should we be scapegoats for all that has gone wrong. On the one hand, discoveries have often been made by scientists for whom the practical applications — for good or ill — were entirely unexpected. On the other hand, and perhaps more typically, possible types of use were often if dimly foreseen and responsibility for realizing them shared with those who translated scientific advances into practical effect through industrial or military technology. In the debate about ultimate responsibility, we scientists have too often claimed credit for what has been beneficial while rejecting blame for what has not. It would not be helpful to attempt a detailed score sheet of credit and blame, but it is important to establish that science and scientists share responsibility to varying extents from one instance to another, not only for the gains, but also for the losses that 'progress' has brought to mankind.

Science, for example, has contributed massively to human misery through seeding a process of dismantling of social

cohesion. Whatever constraints on self-expression are suffered by individuals within traditional stable societies, they are held in a web of relationships which they support and by which they are supported; and those societies could often have ordered governance even when communities were materially very poor. It seems clear that these patterns are breaking up — but to be replaced by what? The paradigms are shanty towns of cities in the poorer countries; slums for illegal immigrants in the United States; and tower-block and housing-estate cultures harbouring delinquency in Britain. The driving forces for these developments are socioeconomic — shifts and changes in population, trade, travel, transport and the like. These have germinated from steps in scientific progress linked with technological innovation which, like acorns planted beneath a building, have insisted on growth towards the light at whatever cost to the stability of the edifices above. The working through of the consequences of scientific discovery seems to have a blind and automatic logic of its own which, once set in train, follows its own remorseless progress regardless of consequences.

The scientific revolution has also sent out ripples, whirlpools and hurricanes to disturb, destroy and swallow up many parts of traditional belief systems that have evolved over millennia as coping strategies to protect social order and function in the face of plagues, famines and disasters. Some belief systems have resisted the pressure to accommodate to the scientific world-view, for example various forms of Christian and non-Christian fundamentalism. Whether in the developed or developing world, these movements have sought to survive unchanged by retreating into closed or partly closed communities. Others, mostly within the mainstream of Western Christian traditions, have explored possibilities for intellectual accommodation, only to discover that the costs are disarray and compromise — and that their reformulation in terms of modern rather than historical mythologies is neither easily nor quickly achieved.

For sanity's sake, societies must now evolve belief systems to incorporate the scientific dimension. But precious little constructive help seems to be forthcoming from the scientific community. Galileo and Darwin agonized about the implica-

tions of their conclusions on their societies' belief systems. Not so many in the scientific community of today, the most vocal representatives of which appear to take glee in the intellectual vandalism of scoring easy and unhelpful points to deny forms of truth other than the scientific, and to equate religion with the useless and superstitious ornamentation of life. There are those on the other side who deplore the rise of science and talk down its continuing advances, and who must share equal responsibility for the stand off; but how can we expect society to acquire more sympathy for science when we ourselves show such little inclination to understand society?

The lack of an adequate and nutritious food supply, the burden of the great infectious diseases, and the population explosion have long been recognized as three major blights on the future of the majority of mankind. Enormous effort and idealism have been committed by the scientific community to the first two of these three great problem areas — to bring about the green revolution; and in the quest, which must surely succeed in the next decade or two, for cheap and effective vaccines against parasitic and other infections.

But no similar scientific bandwagon has yet rolled for the population issue as it needs to be addressed on behalf of poorer countries. Why not? Scientists could blame the lack of political will to address the problem, but this is not convincing. Since when, except when its own welfare and livelihood has been at stake, has it been typical of the scientific community to need or heed the call of political imperative to rally to address an important issue? The truth is that the topic has simply just not caught on. Scientific fields are initiated by the opportunism, enthusiasm and specific insight of gifted scientific leaders, who create frameworks to be filled in and extended by followers and disciples. Whether or not these enterprises bear any relation to social or other extra-scientific priorities has usually — subject to a minor qualification to which I shall return later — been a matter of pure chance.

To summarize, science has assaulted society on two fronts by commission and on another by omission: through technology, commerce and politics on the physical conditions in which people live; through ideas and interpretations on the

MISSIONS PROPOSED FOR A NEW RESEARCH COUNCIL SYSTEM FOR THE UNITED KINGDOM

Research councils should be dedicated to mission defined by distinct and coherent issues central to human life, specifically:

The planet Earth, its microscopic constitution and place in the Universe
 The biosphere, its health and productivity
 The planet Earth, its condition and its physical resources
 Human health
 Products and processes
 Human and social potential

Each mission should integrate fundamental scientific research in all relevant disciplines (whether mathematical, physical, biological and/or sociological) with problem solving and applications for economic, social and cultural benefits.

dreams, myths and beliefs about life and the after life which contribute to spiritual resources; and through the priorities for scientific enquiry having failed sufficiently to parallel social needs. Though the first of these onslaughts has been mediated by others, the last two have flowed directly from science and scientists rather than caused by the hand of anyone in business, politics, engineering or the military. Is there any wonder that we are regarded as a malevolent force?

Scientists themselves frequently insist that the necessary and sufficient condition for productive and beneficial research is merely that it should engage the intellectual interests and curiosity of individual investigators who are creatively gifted; and that on this basis alone future investments in scientific activity should be made.

The 'curiosity motive', a powerful and effective human driving force, has been analysed by Anthony Storr in terms of the psychodynamics of the personality which impel men and women to artistic as well as scientific creativity¹. When we create something ourselves or recognize the creations of others, we are attempting to integrate and reorganize our own experience. Such forms of creative expression are as necessary to our development, individually as persons and collectively as a species, as were cave paintings to early man. Individual anxieties may then shape and add urgency to this general drive; Einstein and Newton, for example, may have been driven by circumstances of early life to develop their own inner worlds of creative thought, to provide explanatory schemes that would alleviate the discomforts of what was to each of them an arbitrary and contradictory world.

When even the most individualistic and antisocial among us has some need for social experience and social approval, it would not be surprising if, deep down, even the most 'curiosity driven' scientist did not hanker to contribute through his work to human progress. This is an argument developed by Jacob Bronowski who has asserted², in a remarkable passage written in 1953 anticipating the deve-

lopment of information technology, that:

Newton turned naturally to astronomy because it was the subject of his day, and it was so because finding one's way at sea had long been a practical preoccupation of the society into which he was born. Faraday worked all his life to link electricity to magnetism because this was the glittering problem of his day; it was so because his society, like ours, was on the look-out for new sources of power. Consider a more modest example today: the new mathematical methods of automatic control, a subject sometimes called cybernetics, have been developed because this is a time when communications and control have in effect become forms of power.

These inventions — optics, electricity and cybernetics — have been directed by social needs and they are useful inventions; yet it was not their usefulness which dominated and set light to the minds who made them.

I have noticed that scientists working in medical research, even on the most abstract problems remote from any application imagined at present, can draw powerful motivation from the overall remit to contribute to the improvement of human health.

To argue from Newton, Faraday or Einstein may be to exalt the average scientist — or even one that reaches the level of conventional eminence — far above his or her station. Like clergymen in the Barchester novels of Anthony Trollope, who were by no means all sinless men of God but often, on the contrary, had to contend with motivations of a trivial or base nature, we scientists may have more than an eye on the main chance and, especially nowadays, the threat to our financial support. Represented more or less strongly in many of our psyches are 'stamp collectors' amassing publications, careerists who want jobs for life with regular promotion, tradesmen prepared to sell services, scholarship pupils who want to shine at the top of the next class, escapists who take comfort from an illusion of science untouched by original sin, would-be gurus and empire builders who appropriate for science areas beyond its sphere. Our motives are no more pure than those of business people, politicians, bureaucrats or the military whom we often

blame for the misuse of science.

Repeated disclaimers from the scientific community of any share in responsibility for the practical and social consequences of research in the past, and continued insistence that our contract with society for future funding should take the form of a blank cheque (or a proportion of Gross Domestic Product) to follow our interests and satisfy our scientific curiosities, reveal a self-centredness and selfishness that is easily seen through from the outside.

None of this implies that all scientists should divert their attention from hands-on science to social and political issues. Like other modern communities, that of scientists needs a differentiation of function between members. It is at the level of leaders and policy makers of science that we need massive adjustment towards objectivity and realism. Unless we begin to see ourselves as others see us, we cannot begin to relate to others in the way that is now so important for us and, did they but know it, for them.

My conclusion is a simple proposition behind which, were enough scientists to unite, we could take a major step towards a new contract and more sympathetic working with society. This is that science does not exist solely for science's sake, but is inseparable from wider missions which can be formulated so that the public will understand and accept them as crucial to our common future. For example, the scientific community in Britain should seize the opportunity provided by the Minister for Science's plans to publish a policy document for the organization of the scientific enterprise, by volunteering for a role and function defined by such missions. Indeed, my own research council has announced proposals for one way in which the British system could be re-organized along these lines (see box).

This view could represent a starting point for an effective public debate, to establish the relationships between past achievements and future goals as seen by the scientific community; the agenda of problems and priorities as seen from within society; those products desired by society that science expects to deliver and those that might be found difficult; and the unsolicited surprises that may yet be in store from future advances. Beyond the practical and material dimensions, we could address together the wider questions of ethics and beliefs. The ultimate prize for which we would work is a science that is integrated into society. □

D. A. Rees is Secretary and Chief Executive of the Medical Research Council, 20 Park Crescent, London W1N 4AL, UK.

1. Storr, A. *The Dynamics of Creation* (Martin, Secker and Warburg, London, 1972).

2. Bronowski, J. *Science and Human Values* (Harper, New York, 1959).

How and why of vesicle formation

Much has been done in the past few years to throw light on the energetics of vesicle formation in monolayer and bilayer membranes. The physics of the process is becoming clear, but the chemistry remains obscure.

THE justifiable complaint that mainstream molecular biology is too neglectful of the quantitative side of the scientific process cannot, mercifully, be levelled at the whole of biology. On the contrary, there are some niches in which the application of techniques familiar in physics to problems in biology is well on the way to being a growth industry. One of the more spectacular of these is the study of the budding of vesicles from lipid membranes, which can be traced back at least as far as an article by P.B. Canham in 1970 (*J. Theor. Biol.* **26**, 61; 1970); the objective, then, was simply to show that if, for example, a closed membrane forms a vesicle-shaped bud, the explanation must be that the budded form has less free energy than the unbudded form.

Nobody will dispute the importance of the process of vesicle formation. Cells that secrete enzymes or hormones do so by means of vesicles budding from their surfaces. Neurotransmitters are handled in the same way, as is the transport of cellular components within a cell, between the successive layers in the Golgi apparatus, for example. Replicated virus particles similarly emerge from infected cells by a budding process, while the inverse of that process, endocytosis, is the way in which cells ingest all but the simplest molecular structures from the outside. And what is to be made of the humble red blood cell, which is simply a free-standing membrane filled with haemoglobin whose normal shape is that of a doughnut or bagel?

And on the face of things, the budding of vesicles from the surface of a cell should be an easy problem. Just think of the outer membrane as a simple homogenous bipolar monolayer and suppose that the departure of the membrane from perfect flatness entails an energy cost — a 'bending energy' — whereupon the energy of the membrane will be simply a function of the bending energy and the curvature over the surface. (The analogy with the stability of a drop of, say, water is not exact.)

In practice, there are complications. The strict analogue of surface tension is the relatively smaller force that holds a membrane laterally together. The bending energy presumably arises differently, by forcing on the molecules that constitute the membrane an orientation different from that in a flat membrane, perhaps bringing similarly charged groups closer together. Moreover, there is no reason why the forces in such a membrane should not be asymmetrical; bending in one direction may require more

energy than bending by the same amount in the other. Symmetrical bilayers will naturally be different.

In the end, the calculation of the shape of a cell, and by extension of the vesicles that bud from it, is a matter of finding the closed surface whose energy is microscopically defined by its radii of curvature (in principle, there are two) at every point and two parameters representing the bending energy. All this has been elegantly explained in a review by Reinhard Lipowsky (*Nature* **349**, 475–481; 1991).

Among other things, this argument explains why vesicle formation is such a common part of life. Start with a flat membrane and there are two components of the energy: the bending energy, zero for a symmetrical membrane, and the peripheral energy arising from the interaction of the edge of the membrane with the outside medium. That energy must be positive, perhaps because of hydrophobic interactions, or the membrane would not have formed in the first place. But the bending energy of a spherical closed membrane turns out not to depend on the radius of the sphere that it forms, but only on the numerical values of the two bending energy parameters.

So if the dimensions of the membrane are sufficiently large, it must be energetically favourable that the membrane should wrap up into a sphere. But, 'other things being equal', the energy embodied in a large sphere should be less than that of two smaller spheres of the same surface area. Endocytosis should be favoured, the budding of vesicles not. But that is not what real life shows, not to mention the increasingly elegant experiments with synthetic membranes that people have devised.

What can be the explanation? That 'other things' are not equal, of course. Predictably, the burgeoning theory of the membrane-budding industry is busily relaxing the assumptions underlying the simple calculations. There are several variables to play with. Experiments show that temperature can affect the configuration of a cell or a synthetic membrane. (Because membranes expand more rapidly than their contents, higher temperatures favour vesicles.) The difference of osmotic and mechanical pressure across the surface should matter, although very little has been done with that.

In some real-life membranes, it may also be significant that some of the molecular components are linked laterally to others by covalent forces; a year ago, B. Fourcade from the Institut Laue-Langevin at Grenoble

and M. Mutz and D. Bensimon from the Ecole Normale Supérieure in Paris used a synthetic membrane whose components were sideways polymerizable to recover toroidal vesicles closely resembling erythrocytes (*Phys. Rev. Lett.* **68**, 2,551; 1992).

A further complication is that the quantity that truly determines the shape of a membrane surface is the free energy, which depends on entropy as well as energy. People are busily measuring the fluctuations of membrane surfaces to make some estimate of how important that may be. Horrendous problems in topology and the variational calculus arise if people are not careful.

Another assumption to be relaxed is that the chemical composition of a membrane is uniform, which is plainly far from the truth. One of the functions of the cell membrane, after all, is to carry the receptors and other proteins that allow a cell to interact with others. This topic has been enlivened in the past few years by the discovery that, in membranes made of mixtures of phospholipids and cholesterol, two separate phases, both of them fluid, can be seen to separate. One is rich in phospholipids, the other in cholesterol.

Two complementary aspects of bilayer membranes with more than one component have now been explored. A few weeks ago, Udo Seifert from the Solid-State Physics Institute at Jülich in Germany showed how the energetics of the problem would allow embryo vesicle buds, perhaps induced by temperature change, to induce inhomogeneities of membrane composition, perhaps forming patches of similar membrane molecules around a nucleating centre.

Now Frank Jülicher and Reinhard Lipowsky (see above) from the same institute have turned the problem around, and have shown that if there is some degree of phase separation on the membrane, the interaction energy along the contour separating the two domains may be just enough to pinch off the membrane carrying one of the phases as a separate vesicle. It is remarkable that there are experimental data to show this to be the case.

None of this implies that it will soon be possible to calculate the production of vesicles carrying, say, neurotransmitter molecules from the synapses of neurons. But it is something to be proud of that the outlines of the physics of the problem have been clarified. What seems most needed now is the injection of a little real-life chemistry into an intriguing line of enquiry.

John Maddox

Getting beyond numerology

William L. Ellsworth

It sounds suspiciously like a question from an aptitude test: given the sequence of numbers 24, 20, 21, 12 and 32, what's the next value? If the next number has to be at least 27, how does that change your answer? In the absence of anything else to go on, one can readily imagine a host of approaches to this problem, ranging from numerology to Poisson statistics.

Indeed, many approaches have been tried to predict the next number, as this particular sequence is the list of intervals, in years, between the known ruptures of the San Andreas fault at Parkfield, California (Fig. 1). Earthquakes with nearly identical magnitudes ($M=6$) appear to have ruptured the same segment of the San Andreas fault in 1881, 1901, 1922, 1934 and 1966. Large foreshocks of the great Fort Tejon earthquake of 1857 have also been interpreted as being part of this sequence. Whether they are more than fortuitous random numbers, however, continues to preoccupy seismologists^{1,2}.

The standard model interprets San-Andreas-type earthquakes as part of a cycle of elastic strain accumulation within the brittle crust and release by slip on faults³. Slip in the earthquake lowers the stress and begins the next cycle. The jerky, stick-slip motion of a block being pulled by a spring is the usual analogue. In such a zero-dimensional model, the block slips forward only when sufficient force has built-up in the spring. Steady extension of the spring establishes a sequence of slip events analogous to repetitive slip along one fault segment.

The apparent regularity of $M=6$ earthquakes at Parkfield (22 ± 6 years), coupled with their rupture of a confined, 30–40-km segment of the fault, led to

an optimistic interpretation of the sequence^{4,5} that foresaw a 95 per cent chance of the next event by the end of 1992. Although the rigorous statistics of the sequence have 95 per cent confidence bounds of ± 18.6 years⁶, it was argued that the 1934 earthquake was triggered early, with the 1966 event falling back on a more narrowly circumscribed time line⁵. Now that, in 1993, this hypothesis can be set aside, geophysicists are refocusing efforts to understand the whole seismic cycle at Parkfield, and the processes that control nucleation of the characteristic earthquake (see box).

Two papers shed new light on this question by raising the level of sophistication in the modelling of Parkfield fault behaviour¹, and by making more critical comparisons between the two most recent earthquakes².

Yehuda Ben-Zion, Jim Rice and Renata Dmowska of Harvard University have rethought¹ the loading cycle for the Parkfield segment, by explicitly considering the influence of great earthquakes on the adjoining part of the fault, such as last occurred in 1857. In their viscoelastic model of the fault, a weak lower crust underlies the elastic, seismogenic upper crust. The whole system is loaded by imposing a far-field plate velocity on the boundary of their three-dimensional finite element model.

When an 1857-type event occurs, it transfers an elastic load to the Parkfield segment and the surrounding crust. With time, the stresses in the lower crust relax, introducing additional time-dependent loading of the Parkfield segment. The effect of these perturbations is to increase the rate of model Parkfield earthquakes

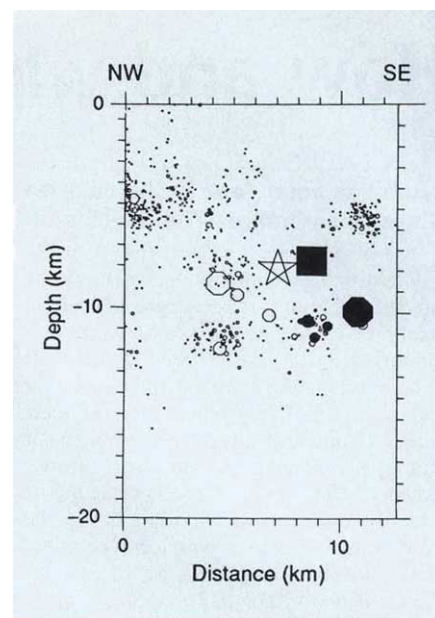


FIG. 2 Cross-section of seismicity since 1980 along the San Andreas fault near the hypocentre of the 1966 Parkfield, California, earthquake (star). Slip zones of earthquakes occurring between 1975 and 1993 are shown, assuming an average stress drop of 3 megapascals (30 bars). Activity in October 1992 seems to have begun the process of eroding the 'asperity' holding the segment (filled octagons). The April 1993 sequence (filled squares) broke an area near the 1966 hypocentre.

for several decades after an 1857 event. After sufficient time has elapsed, the stressing rate at Parkfield falls below the steady-state rate, and Ben-Zion *et al.*'s model predicts a further decline in the rate of characteristic events. Perhaps, because the most recent interval is the longest (32 years) and the open interval (27 years) is second longest, there is some support for this view. As the authors note, the poor quality of the historical record before 1901 leaves no solid evidence of more frequent events in the decades just after the 1857 earthquake.

An additional complication is created by the possibility of spatial variations in fault slip from one Parkfield earthquake to the next⁷. If there are substantial differences, this alone could account for the irregularity in recurrence intervals.

In both 1934 and 1966, the Parkfield earthquake nucleated at the same point, at the northwestern end of the fault segment, and propagated to the southeast. Triangulation measurements show that the 1966 earthquake ruptured for a total distance of 35–40 km, while the 1934 earthquake was 10–15 km shorter. If the two events were substantially different at the north end, where rupture started, this could have a considerable impact on recurrence estimates and forecasts of the next event.

Unfortunately, the geodetic data at the north end are sparse, and no common

FIG. 1 Centre of attraction. The old Town Hall at Parkfield, California, stands as a monument to the pioneers who founded this tiny hamlet (population 34) beside the San Andreas Fault, now the most closely watched fault segment in the world. Three times since its construction in 1882 the building has undergone unplanned modification necessitated by characteristic Parkfield earthquakes¹³. Following the 1901 earthquake the brick chimney top was replaced with a pipe; the second storey was removed after damage suffered in the 1934 earthquake; and in 1966 the badly damaged fireplace was removed. (Photograph by J. Langbein.)



observations span both events. Paul Segall of the US Geological Survey and Stanford University and Yijun Du of Stanford University were inspired to take a new look² at this question, because the initial portions of the 1934 and 1966 seismograms are nearly identical⁴, implying a very similar start for both earthquakes. They attacked this problem by determining the permissible range of differences between the two events using the technique of generalized cross validation. Their approach was to find models of slip on the fault plane that minimized the difference between the earthquakes, while optimally predicting each data point when it is omitted from the modelling.

When they do not penalize dissimilarity between the two shocks, Segall and Du find complementary slip distributions, with large slip in 1934 at the northwest, and large slip in 1966 to the southeast. If true, it would help reconcile the long interval between the events, as larger displacements to the northwest, where the earthquakes start, would require more time for the stress to recover following 1934 than following 1966.

Triangulation data alone, however, do not provide a unique solution. A model for the 1934 event with slip identical to the 1966 event along the northern two-thirds of the rupture is also consistent with the observations. As Segall and Du point out, the average strain released by slip in 1966 is likely to have recovered. So, from the perspective of the strain budget during the seismic cycle, Parkfield appears primed and ready to go. Recovery of the strain released in the last earthquake thus seems to be insufficient, by itself, to bring about the next one.

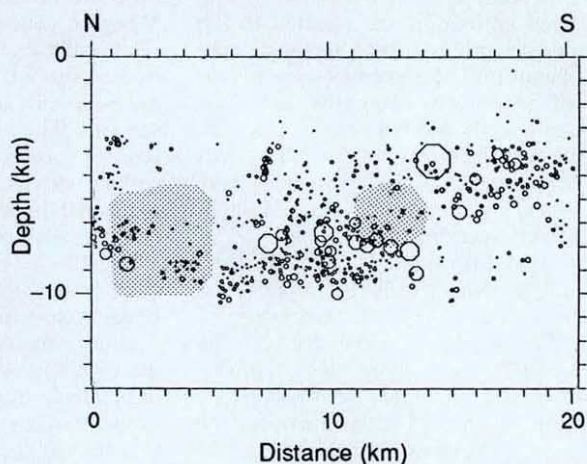
Foreshocks, or small earthquakes occurring within minutes to a few hours of a larger event, and at nearly the same location, are the most obvious manifestations of the triggering process. Both the 1934 and 1966 earthquakes were preceded by $M=5$ foreshocks, and any event of $M \geq 4.5$ near the 1966 hypocentre is viewed as a potential foreshock. Understanding the processes occurring on the fault in the time interval between the foreshock and mainshock is one of the central goals of the Parkfield experiment⁶.

On 20 October 1992 a potential foreshock, $M=4.7$, occurred, triggering a public warning of the next Parkfield earthquake, the first such warning of any imminent earthquake in the United States⁸. The prediction was for a 37 per cent chance of the characteristic earthquake within 72 hours. Nothing happened⁹. The October activity represented the first significant slippage within the proposed hypocentral asperity in more than a decade (Fig. 2). Earlier last month, another part of the asperity failed in an

Characteristic earthquakes

A MODEL of moderate-magnitude San Andreas earthquakes occurring on faults exhibiting both aseismic creep and seismicity is emerging in which the eventual earthquake slip zone remains locked during the strain-accumulation phase of the seismic cycle. Such a locked zone has sometimes been called an asperity.

Cross-section of seismicity along the Calaveras fault in central California. Slip zones of earthquakes occurring between 1975 and 1992 are shown, assuming an average stress drop of 3 megapascal (30 bars). The stippled region in the centre represents the slip zone of the 16 January 1993 Gilroy earthquake ($M=5.2$), inferred from aftershocks and rupture directivity. The stippled region at the left of the cross-section represents slip zone of the 1979 Coyote Lake earthquake ($M=5.9$), as determined by H. Liu and D. V. Helmberger¹². Both preshocks and aftershocks are generally located outside the earthquake slip zone.



Smaller-magnitude earthquakes during this phase generally happen outside the asperity, as is currently observed around the Parkfield nucleation zone⁸.

Such asperities seem to persist through multiple seismic cycles, and break in nearly identical, or characteristic, earthquakes. They also leave a clear signature of themselves in the details of their seismograms, which repeat with remarkable coherency from one recurrence to another.

Is the characteristic earthquake behaviour of the Parkfield segment all that unusual? So few repeats of earthquakes have been identified in the seismological record that special circumstances at Parkfield might be responsible for its repetitive behaviour. This picture is changing, however, as detailed studies of smaller and more numerous earthquakes within the San Andreas fault system are yielding abundant examples of similar characteristic earthquake behaviour.

Oppenheimer *et al.*¹¹ applied the characteristic-earthquake model to the Calaveras fault, and identified two quiet zones on the fault plane surrounded by numerous microearthquakes. They proposed that each zone represented a locked portion of the fault where strain was accumulating. Further, they linked each quiet spot to a historic moderate-sized earthquake. Their southern zone, just east of the town of Gilroy, California, had last ruptured in 1949 in an $M=5.2$ event (see figure, above). A simple comparison of the fault displacement for an event of this magnitude, and the product of the elapsed time and fault slip rate suggested that strain released in the last event should have recovered, and thus a repetition of the event was possible. On 16 January this year, the anticipated event occurred.

W.L.E.

$M=4.4$ event, significantly closer to the 1966 hypocentre. Fortunately for the residents of the region, the earthquake didn't occur this time, either. But no one doubts that it will someday.

As for what to do after it does occur, Clarence Allen of Caltech aptly summed it up in 1986¹⁰. "It should be borne in mind

that following the next Parkfield earthquake, perhaps the most promising place to predict another moderate-sized earthquake will *again* be Parkfield." □

William L. Ellsworth is at the US Geological Survey, Menlo Park, California 94025, USA.

1. Ben-Zion, Y., Rice, J. & Dmowska, R. *J. geophys. Res.* **98**, 2135–2144 (1993).
2. Segall, P. & Du, Y. *J. geophys. Res.* **98**, 4527–4538 (1993).
3. Reid, H. F. *The California Earthquake of April 18, 1906* Vol. 2 (Carnegie Institution, Washington, 1910).
4. Bakun, W. H. & McEvilly, T. V. *J. geophys. Res.* **89**, 3051–3058 (1984).
5. Bakun, W. H. & Lindh, A. G. *Science* **229**, 619–624 (1985).
6. Michael, A. J. & Langbein, J. *EOS* **74**, 145 (1993).
7. Thatcher, W. *Nature* **341**, 432–434 (1989).

8. Parkfield Working Group *EOS* **74**, 152–153 (1993).
9. Finn, R. *Nature* **359**, 761 (1992).
10. Allen, C. R. *et al.* *A Proposed Initiative for Capitalizing on the Parkfield, California, Earthquake Prediction* (National Academy Press, Washington, 1986).
11. Oppenheimer, D. H. *et al.* *J. geophys. Res.* **95**, 8483–8498 (1990).
12. Liu, H. & Helmberger, D. V. *Bull. seismol. Soc. Am.* **73**, 201–218 (1983).
13. Thomason, D. L. *"Cholama" the Beautiful One, Cholama Valley History and Its Pioneer People* (Tabula Rasa, San Luis Obispo, California, 1988).

The power of clonal selection

Klaus Rajewsky

ON page 271 of this issue¹ Lozano and colleagues unveil a new facet of the fascinating process by which cells expressing optimized antibodies are selected in the immune system. By doing so they document beautifully the evolutionary advantage of genetically equipping antibody-producing cells with a single antibody specificity. They also show that cells that have escaped this principle have the means to silence the expression of undesirable specificities later on.

As first postulated by Burnet², antibody-producing cells (B lymphocytes) are predetermined for the expression of antibodies of single specificity. At the genetic level, this is achieved by a principle which is abbreviated below as allelic exclusion: functional genes encoding the variable (V) regions of antibody heavy (H) and light (L) chains are generated through a process of gene rearrangements during B-cell development, and only one functional gene of each class is generated per cell³. The mechanism of allelic exclusion is still not fully understood, but there is strong evidence that it is controlled through the programme of gene rearrangements in a B-cell-autonomous fashion⁴⁻⁷.

Why does the immune system make so much of an effort to produce monospecific B cells? One argument is that this is the most economical way of getting rid of autoaggressive cells through negative selection, while preserving the rest of the antibody repertoire. Another is that the clonal distribution of antibody specificities allows for the most efficient triggering of specific immune responses, through selective activation of only those cells that are selected for the appropriate antibody specificity. The force of these arguments was such that Burnet's theory of clonal selection quickly became the prevailing paradigm in immunology, long before its experimental demonstration.

Advantages

The evolutionary advantage of the clonal distribution of antibody specificities in the attempt to generate specific antibody responses could merely lie in the selective production of antigen-specific antibodies by terminally differentiated, antibody-secreting plasma cells. An additional possibility is that it also optimizes the expansion and selection of B cells in the response.

Lozano *et al.* now provide suggestive evidence for this latter mechanism, from an unexpected angle. In T-cell-dependent antibody responses, antigen-specific B cells undergo rapid and extensive clonal expansion in substructures of the immune

system, the so-called germinal centres. During this process, somatic point mutations are introduced at high rate into the V-region genes of the proliferating cells. Only cells expressing high-affinity antibodies survive and become available as memory cells for secondary antibody responses. The amazing power of this process of positive selection implies that cellular survival depends on signals given to the cells in dependence of their receptor (that is, antibody) specificity (for review see ref. 8).

Lozano *et al.* used transgenic mice harbouring several copies of a rearranged L-chain gene encoding an L chain dominantly expressed in a particular T-cell-dependent anti-hapten response. The group had shown earlier that transgene-encoded L chains participate in this response, and that the transgenes are subject to somatic hypermutation as much as are endogenous V-region genes. They have now further analysed the structure and expression of the various transgene copies present in hybridomas isolated from the transgenic mice after they had been immunized and were secreting high-affinity antibodies carrying transgene-encoded L chains. Because the hypermutation process is random to a large extent within antibody V-region genes, one would expect that in a given cell only one of the multiple copies would have accumulated the particular point mutations leading to high-affinity hapten binding (as defined in earlier work⁹). Would the other copies of the transgene, mutated or unmutated, also be expressed by the cell?

Lozano *et al.* found most of the transgenes harboured by the four hybridomas analysed had undergone somatic mutation and were therefore distinguishable from each other at the messenger RNA level by nucleotide sequence. Furthermore, transgene-encoded L chains could often be distinguished by charge, as predicted from the variations in amino-acid sequence. For any given hybridoma, expression of the transgene turned out to be uneven at the levels of mRNA and/or protein. Translational termination codons had been introduced into some transgenes by the hypermutation process. This might have made the mRNA transcribed from those genes less stable and could have led to their relative underrepresentation in the cells. In other cases, transgenes appeared to be transcriptionally entirely silent, for unknown reasons. Most strikingly, in the two hybridomas in which one transgene copy had accumulated the two point mutations previously shown to confer to the antibody a ten times higher

affinity for the hapten than that of the wild type, the expression of that particular transgene copy by far dominated that of the others.

Two biological messages are inherent in these results. First, antibody-producing cells, if artificially offered the opportunity to express multiple antibody specificities, have at their disposition a variety of mechanisms by which they can modulate the pattern of expression of the corresponding antibody genes. This is also suggested by the gradual counterselection of transgene-expressing cells at the expense of cells expressing endogenous immunoglobulin genes, as is generally seen in immunoglobulin-transgenic mice. Second, in the antibody response there is strong selection for monospecific cells, and cells homogeneously expressing high-affinity antibodies win in the competition. Thus, although the mechanism of this selection process is not yet understood, the evolution of allelic exclusion is of obvious advantage for the organism in terms of the efficiency of the antibody response.

Stringency

More generally, the new results are yet another demonstration of the power of cellular selection operating in the generation of a functional B-cell compartment. The common principle in the various phases of B-cell development is the stringent selection against useless or non-optimal cells — be they cells that have not managed to express an antibody at the surface, that express high-affinity autoreactive antibodies, that fail to generate high-affinity antibodies when undergoing somatic hypermutation in antigen-driven responses, or (as Lozano *et al.* show in the present work and earlier data⁶ had suggested) that have escaped allelic exclusion. (For review of these matters, see ref. 8.) In this sense, the mechanisms by which transgene expression is differentially modulated in the hybridomas of Lozano *et al.* may contribute to clonal selection as such. Moreover, working out these mechanisms may shed light on the problem of the quantitative control of gene expression beyond the cells of the immune system. □

Klaus Rajewsky is at the Institute for Genetics, University of Cologne, D-5000 Cologne 41, Germany.

1. Lozano, F., Rada, C., Jarvis, J. M. & Milstein, C. *Nature* **363**, 271–273 (1993).
2. Burnet, F. M. *The Clonal Selection Theory of Acquired Immunity* (Cambridge University Press, 1959).
3. Tonegawa, S. *Nature* **302**, 575–581 (1983).
4. Alt, F. W. *et al.* *EMBO J.* **3**, 1209–1219 (1984).
5. Reth, M. in *Encyclopedia of Immunology* (eds Roitt, I. M. & Delves, P. J.) 47–49 (Academic, London, 1992).
6. Kitamura, D. & Rajewsky, K. *Nature* **356**, 154–156 (1992).
7. Rajewsky, K., Ehlich, A., Müller, W., Taki, S. & Kitamura, D. *Nature* **359**, 371–372 (1992).
8. Rajewsky, K. *Curr. Opin. Immun.* **4**, 171–176 (1992).
9. Berek, C. & Milstein, C. *Immun. Rev.* **96**, 23–41 (1987).

Process and production

I. Colin Prentice

PHOTOSYNTHESIS by terrestrial vegetation extracts about 120 billion tonnes of carbon from the atmosphere annually — that is over 20 times more than is released by fossil-fuel burning, or a sixth of all the carbon in today's atmosphere¹. Net primary production is the fraction, about half, of this carbon flux that goes into plant tissues, and we need to know how it will change as levels of CO₂ rise and the Earth warms. On page 234 of this issue², Jerry Melillo and others provide an answer, based on a computer model of the

tem in the TEM; a more exact biochemical model is available⁵ and might give a different answer. Also, at high latitudes the TEM shows net primary production increasing with temperature. Melillo's group have demonstrated elsewhere⁶ that this effect depends on the coupling between carbon and nitrogen fluxes. Boreal net primary production is limited by how fast decomposition can supply inorganic nitrogen. The TEM predicts that warming would stimulate decomposition and increase net primary production; without

is not located in the oceans according to current understanding of ocean-atmosphere exchanges¹⁰. Could a CO₂-driven rise in net primary production account for the missing sink? Of course, an increase in net primary production is not the same thing as a net carbon sink. Globally, annual net primary production is almost balanced by heterotrophic respiration, only a small fraction of carbon being sequestered in plant biomass or humus¹. But as an insightful analysis⁹ has suggested, any increase in net primary production over a time scale of years or decades would be expected to produce a (transient) net carbon sink, so the results of Melillo *et al.* might point in this direction.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Moist tropical forest, which would be responsible for a large part of the total response of the biosphere to doubling in CO₂.

underlying biophysical and biochemical processes. They conclude that if CO₂ doubles and climate changes as projected by climate models, global net primary production could increase by 20–26 per cent, partly because of CO₂-induced 'fertilization' of plant growth in the tropics and partly because of warmth-induced stimulation of nitrogen cycling in high latitudes.

Measurements of annual carbon fluxes in world ecosystems have been made since the 1970s, but the only predictive models have been simple regressions³ that would probably fail in a changed climate. In the past few years better models have appeared, ones based on the underlying processes of photosynthesis, respiration, transpiration, nitrogen cycling and their interactions⁴. The terrestrial ecosystem model (TEM) used by Melillo and colleagues is the first of these process-based terrestrial biogeochemical models (TBM) to be applied worldwide.

The next phase of research with TBMs is likely to include comparisons between them to assess their robustness. Some of the findings of Melillo *et al.* will certainly stimulate such comparisons. For example, their model suggests that a doubling of CO₂ alone would increase annual net primary production by 8 petagrams (10¹⁵ g) of carbon. This conclusion depends partly on the relationship between photosynthetic rate and the within-leaf CO₂ concentration, which is approxima-

ted in the TEM; a more exact biochemical model is available⁵ and might give a different answer. Also, at high latitudes the TEM shows net primary production increasing with temperature. Melillo's group have demonstrated elsewhere⁶ that this effect depends on the coupling between carbon and nitrogen fluxes. Boreal net primary production is limited by how fast decomposition can supply inorganic nitrogen. The TEM predicts that warming would stimulate decomposition and increase net primary production; without

the nitrogen limitation, warming would merely increase respiration rates and decrease production⁶. So the direction of change could depend on just how the respiration-temperature relationship and carbon-nitrogen coupling are formulated. Because physically exact formulations of these processes are not known, comparisons of alternative models are needed: among other things they may help to show which experiments or measurements would help most to reduce uncertainties. The new results² are offered as a sensitivity experiment. They portray a hypothetical situation in which the present distribution of vegetation co-exists with a changed climate. But over the period needed to double CO₂, vegetation will alter in response to changes in climate^{2,7,8}, and to changing land use by an increasing population. These latter effects could be crucial — for example, the TEM indicates that moist tropical forests are responsible not only for a large part of terrestrial net primary production, but also for a large part of the total biospheric response to CO₂ doubling. If this is so, then large-scale conversion of tropical forests to agriculture would tend to reduce net primary production and to limit the biosphere's further capacity to absorb excess CO₂⁹.

Changes in the biosphere induced by CO₂ may already be happening. A part of the CO₂ released by burning of fossil fuel and by deforestation disappears each year into a 'missing' carbon sink, one which

Indeed, TBMs may prove most useful in helping us to understand the dynamics of the global carbon cycle on short time scales. For example, it would be a breakthrough if a TBM coupled to an ocean biogeochemistry model could be shown to generate the correct seasonal and latitudinal source-sink distribution implied by recent measurements of atmospheric CO₂^{9,10}.

By the time CO₂ has doubled, the carbon balance of the biosphere may well be dominated by other effects — including changes in biome distribution — that are not dealt with by TBMs. Biome models⁷ show that CO₂-induced climate change would alter the global distribution of potential natural vegetation^{8,11,12}. Studies

- Schlesinger, W. H. *Biogeochemistry: An Analysis of Global Change* (Academic, New York, 1991).
- Melillo, J. M. *et al.* *Nature* **363**, 234–240 (1993).
- Lieth, H. in *Primary Productivity of the Biosphere* (eds Lieth, H. & Whittaker, R. H.) 237–263 (Springer, New York, 1975).
- Running, S. W. & Gower, S. T. *Tree Physiol.* **9**, 147 (1991).
- Farquhar, G. D., von Caemmerer, S. & Berry, J. S. *Planta* **149**, 78–90 (1980).
- McGuire, A. D. *et al.* *Glob. Biogeochem. Cycles* **6**, 101–124 (1992).
- Prentice, I. C. *et al.* *J. Biogeogr.* **19**, 117–134 (1992).
- Smith, T. M., Leemans, R. & Shugart, H. H. *Clim. Change* **21**, 367–384 (1992).
- Taylor, J. & Lloyd, J. *Aust. J. Bot.* **40**, 407–418 (1992).
- Post, W. M. *Am. Sci.* **78**, 310–326 (1990).
- Prentice, I. C. & Sykes, M. in *Biotic Feedbacks in the Global Climate System: Will the Warming Speed the Warming?* (ed. Woodwell, G. M.) (Oxford University Press, New York, in the press).
- Prentice, K. C. & Fung, I. Y. *Nature* **346**, 48–50 (1990).
- King, G. A. & Neilson, R. P. *Wat. Air. & Soil Pollut.* **65**, 365–383 (1992).
- Smith, T. M. & Shugart, H. H. *Nature* **361**, 523 (1993).

with different climate and biome models agree on the broad features of the change, notably a major poleward shift of temperate forest types and an expansion of boreal forests into the tundra. The studies vary in how far this forest expansion is offset by the conversion of temperate forests in more continental climates to grassland or shrubland. All show an increase in the potential area of tropical forest, though this is slight in recent, process-based analyses¹¹. Carbon storage at equilibrium tends to increase^{8,11,12}, but the first attempts to simulate the kinetics involved suggest that for decades climate-driven vegetation changes would produce a source, not a sink, for carbon^{13,14}. This is simply because forest decline and oxidation of soil carbon in some regions would occur faster than new forest growth and

accumulation of soil carbon elsewhere. We do not yet have realistic estimates of how large this source would be, or how it would interact with the carbon sink due to increasing net primary production.

Considering all the relevant processes, it becomes clear that the biosphere can produce both negative and positive feedbacks to CO₂-induced global warming, and that the direction of these feedbacks can change with time. To understand the interactions among these processes, in a quantitative and predictive sense, is the difficult (but essential) task for global ecology. □

I. Colin Prentice is in the Global Systems Group, Department of Ecology, Lund University, Östra Vallgatan 14, S-223 61 Lund, Sweden.

OCEANOGRAPHY

Carbon overconsumption

J. R. Toggweiler

A FUNDAMENTAL observation in biological and chemical oceanography is the ocean-wide consistency in the elemental composition of marine organic matter (plankton biomass and undecomposed detrital particles). Bulk organic matter sieved from sea water combines carbon, nitrogen and phosphorus, on average, in the molar proportions C:N:P = 106:16:1 (ref. 1). Dissolved nitrate, phosphate, inorganic carbon and oxygen in the thermocline and deep sea follow the same ratios, suggesting that detrital organic matter oxidized in the interior of the ocean is identical in composition to the organic matter produced at the surface^{2,3}. These constant proportions — known in the trade as Redfield ratios — are often used to link the production of new organic matter to the uptake or supply of nitrate and phosphate.

But measurements by Sambrotto and colleagues on page 248 of this issue⁴ show matters to be more complicated. The authors have tracked the concurrent drawdown of dissolved inorganic carbon (DIC), nitrate and phosphate during spring phytoplankton blooms. They find that the amount of carbon removed from the water significantly exceeds the amount expected based on nitrate/phosphate removal and the C:N or C:P ratio in bulk organic matter. Some 30–40 per cent of the carbon removed from the water cycles through the upper ocean in an unknown way.

Monitoring the simultaneous changes in DIC and nitrate during a phytoplankton bloom may sound like a simple thing to do, but it could not have been done in any meaningful way until recently. The day-to-day changes in DIC reviewed by Sam-

brotto *et al.* amount to only a few micromoles per litre of sea water. Measurements accurate enough to monitor these changes have been possible only since the mid-1980s. Furthermore, the conversion of DIC into organic matter is only one of a number of processes affecting DIC. CO₂ enters the ocean from the atmosphere during spring blooms, raising DIC levels. DIC is also taken up into nonorganic biogenic material, such as CaCO₃ skeletons. Isolating the effect of organic carbon production from these competing processes requires simultaneous high-precision measurements of dissolved CO₂(g) and alkalinity.

Sambrotto *et al.* compile time-series data from the southeast Bering Sea continental shelf, the Bransfield Strait area adjacent to the Antarctic Peninsula, the Grand Banks, and the site of the JGOFS (Joint Global Ocean Flux Study) North Atlantic bloom experiment. In each of these areas, late-winter nitrate and phosphate levels are relatively high, providing a nutrient feedstock for rapid phytoplankton growth. Within a month or so of the beginning of the spring bloom, nitrate levels have typically declined by 3–10 µmoles per litre over the upper 30–40 m, while DIC levels have decreased by 30–130 µmoles per litre. Sambrotto *et al.* conclude that the amount of carbon removed from the water during phytoplankton blooms appears to be 10–14 times the amount of nitrate removed. This is 40–80 per cent greater than the carbon uptake explained by the production of biomass. There must be a large pool of organic material that cycles through the system with an anomalous C:N ratio.

Sambrotto *et al.* attempt to explain

Organic trouble

The difficulty of accurately measuring the traces of organic material residing in sea water was highlighted in a recent issue of *Marine Chemistry*, built on a meeting on dissolved organic carbon (DOC) and nitrogen (DON) held in Seattle in 1991. In one article¹⁰, Y. Suzuki retracted results that had stimulated a lot of excitement when they first appeared in 1985¹¹ and 1988¹² (see my previous News and Views article¹³). In the earlier papers, Suzuki and colleagues unveiled a new technique, high-temperature catalytic oxidation (HTCO), which apparently showed DOC and DON levels to be as much as 2–4 times higher than previously believed. The measurements seemed to highlight an unknown, large, long-lived component of dissolved organic matter that is resistant to microbial oxidation. Also, a strong correlation of the new DOC levels with dissolved oxygen levels in the thermocline suggested that DOC and DON must have a large and unexpected role in the cycling of carbon and nutrients in the ocean.

But now it seems that the actual DOC levels are lower than originally reported by Suzuki and colleagues. Much of the discrepancy seems to come from instrumental carbon contamination, or 'blanks', perhaps associated with the HTCO catalyst¹⁴. (Typical HTCO samples consist of only 100 µl of sea water; undetected blanks at a level of 5×10^{-9} moles can skew results by as much as 100 per cent.) When a large blank is subtracted from HTCO measurements the mysterious refractory DOC pool largely disappears.

The situation is far from bleak, however. Even with the blank correction, HTCO may still extract DOC that older methods do not extract¹⁵. It also seems to offer an inherent precision not seen before. Let's hope the bugs can be worked out. J.R.T.

these results as a consequence of preferential recycling of nitrogen and phosphorus during the decomposition of detrital material. Nitrogen- and phosphorus-rich organic compounds (proteins, amino acids, nucleic acids, ATP) are thought to be consumed by microbial organisms in preference to carbon-rich compounds (such as carbohydrates and lipids). In one model, detrital material is assumed to sink slowly as it is being decomposed. Preferential recycling allows nitrogen and phosphorus to be stripped from the sinking particles in the euphotic (sunlit) zone while a carbon-rich residue settles into the aphotic layers below. The extra nitrogen and phosphorus retained in the euphotic zone fuel still more production and amplify the production of more carbon-rich particles.

There are two problems with this ex-

planation. Simultaneous measurements of organic carbon and nitrogen in marine particulate matter generally do not show a systematic bias towards high C:N ratios either in or below the euphotic zone^{5,6}. (Some evidence exists for high C:P ratios in large fast-sinking particles, but there are methodological problems associated with these observations⁶.) The idea of a continuously modified sinking particle pool runs into severe problems in the deep ocean. The longer and deeper these modified particles sink, the more carbon-rich they would become. This should lead to strongly skewed ratios in the variations in O₂:N, C:N, O₂:P and C:P when these particles are ultimately decomposed in deep water. A study of water below 400 m in the South Atlantic, Indian and Pacific Oceans shows no such modification³.

A second example of preferential recycling calls for the build up of a carbon-rich detrital pool that does not sink. For this scheme to work, a large standing stock of organic carbon must accumulate in the euphotic zone over the course of a bloom, around 15–50 $\mu\text{moles C per litre}$. This amount greatly exceeds the standing stock of phytoplankton, zooplankton and filterable detrital particles (which is less than 5 $\mu\text{moles C per litre}$) but it is not particularly large in relation to upper-ocean pools of dissolved organic material which contain of the order of 100 $\mu\text{moles C per litre}$. Thus, one can construct a model in which dissolved organic carbon (DOC) accumulates through the bloom and decomposes later, perhaps during the summer, or perhaps after mixing into the thermocline the following winter. One should expect to see DOC concentrations rise and fall by around 15–50 $\mu\text{moles per litre}$ with little concurrent variation in levels of dissolved organic nitrogen (DON). Given all the methodological problems associated with DOC and DON measurements (see box), it is too early to tell whether this model works or not.

Recent observations from the Bermuda Station S and the JGOFS Bermuda Atlantic Time Series Station (BATS) widen the

scope of the mystery. The upper ocean around Bermuda is characteristically very low in nutrients. Nonzero nitrate levels, 0.5–1.0 $\mu\text{moles per litre}$, are seen only briefly during mid-winter⁷. Nevertheless, time-series data of surface DIC levels near Bermuda show a regular seasonal cycle with an amplitude of 30 $\mu\text{moles per litre}$ (refs 8, 9). The drawdown phase begins at the winter–spring transition and extends into the summer. Virtually all the DIC drawdown in the spring and summer is driven by the production of organic matter⁸. During this time, there is no nitrate or phosphate in the surface water, and there are very low levels of plankton biomass and very small fluxes of sinking particles⁷. Thus, the mysterious carbon drawdown noted by Sambrotto *et al.* may

exist even in nutrient-poor regions of the ocean where the spring bloom is quite muted.

The anomalous DIC drawdown at Bermuda and at the more productive sites described by Sambrotto *et al.* amounts to 1–3 moles C m⁻². This is a substantial fraction of the annual production of exportable organic matter (new production) at these sites. A large component of the ocean's organic-matter production defies a simple link to nutrient supply. We don't know the form that the anomalous production takes, or where or when it is ultimately oxidized. □

J. R. Toggweiler is in the NOAA Geophysical Fluid Dynamics Laboratory, Princeton University, Princeton, New Jersey 08542, USA.

SOLAR SYSTEM

Vestal voyagers unveiled

William B. McKinnon

VESTA, brightest and third-largest of the asteroids, and named after the virgin Roman goddess of the hearth, turns out to have spawned a small family of minor asteroids and, apparently, other offspring in the form of meteorites found on Earth. R. P. Binzel and S. Xu argue this¹ on the basis of family resemblances discovered spectroscopically.

Spectra of Vesta taken since 1970 indicate that it has a basaltic surface². Basalt is the most common volcanic rock on the Earth and other terrestrial planets, and so it is exciting to contemplate familiar igneous activity on a small world beyond the orbit of Mars. Furthermore, basalts are not uncommon among the meteorites. Some of these extraterrestrial basalts are so young, geologically speaking, that they have been interpreted as coming from Mars. Some are from the Moon. Others form a chemically related group, the eucrites (from the Greek for 'easily distinguished'; see figure), that are geologically ancient. Their crystallization ages are coincident with the beginning of Solar System history³, 4.5 billion years ago, and a genetic link between the eucrites and the surface of asteroid Vesta has long been suspected. This hypothesis has endured, despite difficulties in seeing how material from Vesta's orbit could reach the Earth, because no other main-belt asteroid provided a spectral match to the eucrites. Until now.

In a spectroscopic *tour-de-force*, Binzel and Xu have found data that seem finally to establish Vesta as the parent body of the eucrites and of two other related classes of meteorites, indicating in the process how these meteorites may get to the Earth. They took spectra of more than 100 small, main-belt asteroids (each less

than 10 km in diameter) at or near Vesta's orbital distance; 15 of these had previously been dynamically defined to be Vesta-family objects; that is, they share very similar orbital semimajor axes, eccentricities and inclinations.

Despite the violence done to mythology in ascribing offspring to Vesta, from the standpoint of planetary geology a family of smaller asteroids near Vesta's orbit makes sense. The main-belt asteroids exist on intersecting orbits, and are a collisionally evolving population. Their average encounter speed is around 5 km s⁻¹, too fast for the colliding objects to stick and build into a single super-asteroid. This disruptive situation is attributable to gravitational perturbation of the orbits by Jupiter, and it is rather remarkable that even an asteroid as large as Vesta (520-km average diameter) has retained a basaltic crust at all. There is widespread spectroscopic and meteoritic evidence for igneous differentiation and the formation of planet-like iron cores among asteroids of the inner asteroid belt, but this evidence is coupled with inferences that these asteroids have repeatedly been catastrophically ruptured and reassembled, with only remaining lower mantles and cores surviving⁴ (see my previous News and Views item⁵). So at the very least, Vesta should be accompanied by fragments or chips off its surface layers.

Of the 15 members of the Vesta family surveyed by Binzel and Xu, 12 were found to have visual and near-infrared spectra very similar to that of Vesta and the eucrites. Actually, the spectral details are quite revealing. Ten objects (including asteroid Lennon) are essentially spectrally identical to Vesta and the eucrites. Two

1. Redfield, A. C. *et al.* in *The Sea Vol. 2* (ed. Hill, M. N.) 26–77 (Wiley, New York, 1963).
2. Takahashi, T., Broecker, W. S. & Langer, S. J. *geophys. Res.* **90**, 6907–6924 (1985).
3. Anderson, L. thesis, Princeton Univ. (1993).
4. Sambrotto, R. N. *et al.* *Nature* **363**, 248–250 (1993).
5. Copin-Montegut, C. & Copin-Montegut, G. *Deep-Sea Res.* **30**, 31–46 (1983).
6. Bishop, J. K. B. in *Productivity of the Ocean: Present and Past* (eds Berger, W. H., Smetacek, V. S. & Wefer, G.) 117–137 (Wiley, Chichester, 1989).
7. Bermuda Atlantic Time Series Study, Data Report B-1A, (JGOFS Planning Office, Woods Hole Oceanographic Inst., 1991).
8. Keeling, C. D. in *The Global Carbon Cycle* (ed. Heimann, M.) (Springer, Berlin, in the press).
9. Bates, T. *et al.* at *The Oceanography Soc. Annual Meeting* (12–16 April 1993, Seattle).
10. Suzuki, Y. *Mar. Chem.* **41**, 287–288 (1993).
11. Suzuki, Y. *et al.* *Mar. Chem.* **16**, 83–97 (1985).
12. Sugimura, Y. & Suzuki, Y. *Mar. Chem.* **24**, 105–131 (1988).
13. Toggweiler, J. R. *Nature* **334**, 468 (1988).
14. Benner, R. & Strom, M. *Mar. Chem.* **41**, 153–160 (1993).
15. Wangersky, P. J. *Mar. Chem.* **41**, 61–74 (1993).

are slightly different, having a deeper and sharper pyroxene absorption band at a wavelength of 0.9 μm , indicating a more pyroxene-rich composition (and a slight shift in the band centre implies more magnesium and less iron in the rocks). A portion of Vesta's surface is also more pyroxene-rich and magnesian, as inferred from spectra of Vesta taken at different rotational aspects⁶.

Furthermore, there are two meteorite classes (as noted above) related to the eucrites by similarities in their mineralogical compositions and textures and their oxygen isotope ratios. The first is the diogenites. Whereas the eucrites are basalts (magmatic pyroxene-plagioclase rocks), the diogenites are made of nearly pure magnesian orthopyroxene. Most eucrites and diogenites are breccias, composed of rock fragments cemented by pulverized mineral grains. The other related meteorite group comprises the howardites, breccias formed of various proportions of eucritic, diogenitic, and in some cases, chondritic (the most common meteorite type) fragments. This juxtaposition in breccias, one might say, cements the association between all three groups, and indicates that they formed on the same asteroid.

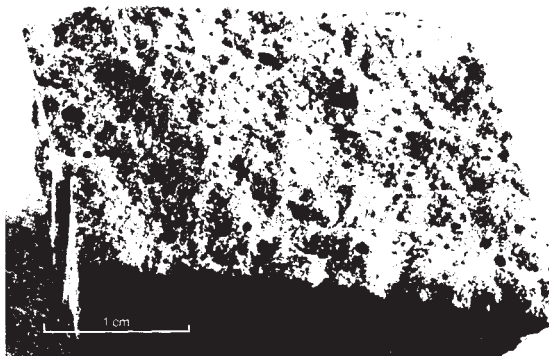
The eucrites are judged to come from surface volcanic material or shallow intrusions, and the diogenites are apparently cumulate rocks that formed at depth. Gaffey's interpretation of the rotational variation in his Vesta spectra was that two large craters (each a few hundred kilometres in extent), products of giant collisions, expose deeper diogenitic material^{6,7}. That two of Binzel and Xu's Vesta chips could also be diogenitic is thus plausible. Proof will come with further near-infrared spectral measurements at longer wavelengths, where the 1.25- μm plagioclase absorption present in spectra of Vesta and the eucrites should naturally be absent.

The other three family members observed by Binzel and Xu possess sufficiently different spectra to be considered interlopers, although one of these small asteroids is apparently olivine rich. An olivine signal is associated with the centre of one of the pyroxene-rich regions of Vesta^{6,7}, possibly due to the excavation of mantle-like material from beneath the crust, and Binzel and Xu note that this one olivine-rich object could also be Vesta ejecta.

The existence of this Vesta family also tells a story about impact physics and the strength of the material. The escape velocity from Vesta, Binzel and Xu estimate, is about 375 m s^{-1} and the 10-km fragments must have survived the collisions that launched them at speeds in

excess of this without disintegrating and dispersing.

It has long been recognized that the collisions that create such fragments in the asteroid belt cannot give them a sufficient kick to reach the Earth's surface. Instead, Jupiter's gravity is invoked again. Asteroids at an orbital radius of 2.5 astronomical units (1 AU is the radius of the Earth's orbit) circle the Sun three times in every jovian year; there is another similar resonance, the v_6 , at 2.04 AU (for low-inclination orbits). The cumulative effect



Piece of the unbrecciated eucrite, Ibitira, which fell in Brazil in 1957. Although shocked, this rock had remained intact since the day it solidified 4.5 billion years ago at the surface of Vesta or a similar asteroid. Most striking are the small vesicles, which resulted from the exsolution of dissolved volatiles (gas) when ascending magma in the asteroid reached the low-pressure, near surface. (From ref. 11.)

of Jupiter's periodic pull is to distort the asteroids' orbit to the point that they cross the Earth's. This takes less than 10^6 years for the 3:1 resonance, and around 10^8 for the v_6 . Wetherill has shown that this mechanism can account for the flux of meteorites at the Earth⁸ and for the population of near-Earth asteroids⁹. But he has long argued⁸ that Vesta, at 2.36 AU, is too far from either resonance to be a significant source of meteorites, the eucrites in particular.

This problem, how Vesta could deliver material to Earth, now appears to be solved. What Binzel and Xu have found are eight non-family (orbitally speaking) asteroids with eucritic or diogenitic spectra, out of the hundred or so examined in the region around Vesta, that occupy orbits extending from that of Vesta out to the edge of the 3:1 resonance. These are clearly associated by their spectra with Vesta, and are probably large surface and subsurface chunks (6–10 km across) blown out to larger semimajor axes by a single great impact. Four of the eight are eucritic and four diogenitic, suggesting relatively deep excavation. Their ejection velocities must have been large, at least 600 m s^{-1} , and it seems clear that multi-kilometre fragments were ejected into the 3:1 resonance, and that kilometre-sized ejecta could have reached

the more distant v_6 (although evidence for the latter is not seen).

Now, Wetherill is certainly right: Vesta should not, in an averaged sense, contribute much material to the Earth. But it got around this, apparently, by suffering a rare impact large enough to have seeded the 3:1 resonance with large fragments. It is the further impact fragmentation of these intermediate parent bodies that provides the eucrites, diogenites and howardites. Three small (1–3-km diameter) near-Earth asteroids, two of which are occasional Earth-crossers, have already been discovered with spectra nearly identical to that of Vesta and the eucrites¹⁰.

To substantiate the Vesta-eucrite connection, it will be necessary to find asteroids with basaltic/pyroxenitic spectra in deeply Earth-crossing orbits (the three near-Earth examples found so far all have perihelia outside the Earth's orbit). Such orbits would explain why there are apparently equal numbers of falls of eucrites and related meteorites in evenings and mornings; fall time is an important constraint on meteorite origins⁸. It would also be comforting to find examples with semimajor axes between 2 and 2.5 AU, which are most likely for asteroids evolving from the 3:1 resonance⁹. The semimajor axes for the three basaltic near-Earth asteroids,

ranging from 1.82 to 2.09 AU, are also consistent with a v_6 source, which Wetherill has long advocated for the eucrites. Finally, Binzel and Xu discuss the difficulties, using present impact theory, of accounting for large fragments ejected at the high velocities indicated by their results. But, as with similar difficulties in explaining the lunar and martian meteorites, nature has clearly imagined a way. □

William B. McKinnon is in the Department of Earth and Planetary Sciences and McDonnell Center for the Space Sciences, Washington University, Saint Louis, Missouri 63130, USA.

1. Binzel, R. P. & Xu, S. *Science* **260**, 186–191 (1993).
2. McCord, T. B., Adams, J. B. & Johnson, T. V. *Science* **168**, 1445–1447 (1970).
3. Dodd, R. T. *Meteorites* 260–262 (Cambridge University Press, 1981).
4. Bell, J. F., Davis, D. R., Hartmann, W. K. & Gaffey, M. J. in *Asteroids II* (eds Binzel, R. P., Gehrels, T. & Matthews, M. S.) 921–945 (University of Arizona Press, Tucson, 1989).
5. McKinnon, W. B. *Nature* **340**, 343–344 (1989).
6. Gaffey, M. J. *Lunar planet. Sci.* **XIV**, 231–232 (1983).
7. McSween, H. Y. *Meteorites and Their Parent Planets* 137–142 (Cambridge University Press, 1987).
8. Wetherill, G. W. *Phil. Trans. R. Soc.* **A323**, 323–337 (1987).
9. Wetherill, G. W. *Icarus* **76**, 1–18 (1988).
10. Cruikshank, D. P. et al. *Icarus* **89**, 1–13 (1991).
11. Wasson, J. T. *Meteorites: Their Record of Early Solar System History* (Freeman, New York, 1985).

Motor neurons find their factors

Anne W. Mudge

In this issue, Henderson and colleagues¹ report results which add to our understanding of how neurons grow, develop and are maintained. As described on page 266, they find that three of the so-called neurotrophins are powerful agents in promoting the survival of rat motor neurons, which, moreover, seem to express receptors for the neurotrophins. The interest in these and related observations is not only scientific, but in the possible development of treatments for a variety of neurodegenerative diseases of motor neurons such as amyotrophic lateral sclerosis.

Like many developing vertebrate neurons, motor neurons in the spinal cord depend on their target cells — in this case, skeletal muscle cells — for their survival. By analogy with the prototypic neuronal survival factor, nerve growth factor (NGF), it was assumed that there would be one or more muscle-derived trophic factors that would specifically support motor neurons during development and, perhaps, throughout life². But what were these putative muscle-derived neurotrophic factors?

The general strategy in the quest to find out was to assay skeletal muscle extracts for their ability to promote survival of enriched embryonic motor neurons *in vitro*. Although survival activity could be demonstrated in this way, no one succeeded in purifying the relevant proteins. Recently, however, success has come from a different route: members of the neurotrophin family have been shown to promote the survival of motor neurons, both during development and following injury, and the paper by Henderson *et al.* is the most recent example of this approach.

Serendipity

In a famous tale of serendipity, NGF was purified from male mouse salivary glands, which, for unknown reasons, contain large amounts of the protein compared to those present in most targets of NGF-dependent neurons. Barde and Thoenen and their colleagues later spent many years purifying a second survival factor from porcine brain that, like NGF, rescues a subset of sensory neurons from naturally occurring cell death; they called this protein brain-derived neurotrophic factor (BDNF) and found that it is structurally related to NGF³. Within months of the publication of the BDNF sequence, six groups independently identified a third member of this newly recognized neurotrophin family, neurotrophin-3 (NT-3), which supports the survival of some sensory neurons, including the propriocep-

tive neurons³. A fourth, more distantly related member of the family was later identified, first in *Xenopus* and then in mammals. It is known as either neurotrophin-4 or -5 (NT-4/5) (see ref. 1).

Although muscle expresses both BDNF and NT-3 messenger RNA, it was thought that neither of these neurotrophins were potential survival factors for motor neurons because they did not support the survival of chick motor neurons *in vitro*⁴. So attention initially centred on the effects of neurotrophin on sensory, sympathetic and brain neurons. In the past year, however, there has been a flurry of papers suggesting that BDNF and NT-3 may indeed serve as trophic factors for motor neurons. Both adult and neonatal motor neurons transport BDNF and NT-3 in retrograde fashion⁵⁻⁷. BDNF rescues substantial numbers of motor neurons that would otherwise die after lesioning of the neonatal sciatic⁶ or facial nerve^{7,8}; in one case, NT-3 also rescued motor neurons, although less effectively than BDNF⁸. Finally, BDNF inhibits the normal cell death of embryonic chick motor neurons *in vivo*⁹.

Using antibodies to identify and purify embryonic rat motor neurons, Henderson and colleagues¹, now show that recombinant BDNF, NT-3 and NT-4/5 are all equally effective at promoting short-term survival of these purified neurons *in vitro*. The dose-response curves for each of the factors are almost identical and are in the picomolar range, implying that each factor acts on its own specific receptor. Further, NT-3 and, at lower levels, BDNF mRNAs are present in skeletal muscle at the time at which motor neuron axons are growing into the muscle and competing for survival factors; the presence of NT-3 mRNA in muscle at this time has also been reported by others¹⁰. Embryonic motor neurons express both trkB (the preferred receptor for BDNF and NT-4/5) and trkC (the preferred receptor for NT-3), but not trkA (the preferred receptor for NGF) (see ref. 11). It seems possible, then, that motor neurons compete for NT-3 or BDNF or both during the period of normal motor neuron cell death.

More detailed studies of the cellular distribution of the neurotrophins, as well as experiments that interfere with their function, are now needed to determine whether these neurotrophins are indeed the elusive trophic factors for motor neurons. Neurotrophins may also be involved in maintaining mature motor neuron integrity and in promoting regeneration after injury. Adult motor neurons continue to express trkB and trkC¹, and BDNF and NT-3 mRNA are expressed in

mature muscle⁷. Moreover, the expression of BDNF is upregulated in denervated muscle whereas that of NT-3 is down-regulated⁷. BDNF expression is also switched on in Schwann cells following peripheral nerve lesion¹².

The neurotrophins seem to act on both sensory and motor neurons, and BDNF and NT-3 are both expressed in skin and muscle, so the old idea that targets would supply a specific trophic factor for each class of projection neuron was too simple. Furthermore, embryonic motor neurons express NT-3 mRNAs transiently^{1,10,13}. Motor neuron-derived NT-3 may support NT-3-dependent proprioceptive sensory neurons that make synapses on the motor neuron. It is also possible, however, that NT-3 may act on immature motor neurons in an autocrine or paracrine fashion, as suggested for sensory neurons that make and respond to BDNF^{13,14}.

Lesion factor

A given cell may well require several signalling molecules for long-term survival^{4,15,16}, and it seems probable that motor neurons do too. Several other defined proteins have been shown to promote motor neuron survival. Of these, the best studied is ciliary neurotrophic factor (CNTF), which was first identified in tissue extracts as an activity that promotes the survival of cultured chick parasympathetic motor neurons, but was eventually purified from mammalian sciatic nerve extracts and then cloned³. CNTF is a cytosolic protein made by myelinating Schwann cells in peripheral nerves and astrocytes in the central nervous system, and it is thought to act as a 'lesion factor' that is released by damaged glial cells¹⁷. It can promote the survival of embryonic spinal-cord motor neurons^{1,2,4} and rescue lesioned or degenerating neonatal motor neurons¹⁸. Whether CNTF is involved in the maintenance of normal neuronal integrity remains to be seen. ►

- Henderson, C. E. *et al.* *Nature* **363**, 266–270 (1993).
- Oppenheim, R. W. *A. Rev. Neurosci.* **14**, 453–501 (1991).
- Thoenen, H. *Trends Neurosci.* **14**, 165–170 (1991).
- Arakawa, Y., Sendtner, M. & Thoenen, H. *J. Neurosci.* **10**, 3507–3515 (1990).
- DiStefano, P. S. *et al.* *Neuron* **8**, 983–993 (1992).
- Yan, Q., Elliot, J. & Snider, W. D. *Nature* **360**, 753–755 (1992).
- Koliatsos, V. E. *et al.* *Neuron* **10**, 359–367 (1993).
- Sendtner, M., Holtmann, B., Kolbeck, R., Thoenen, H. & Barde, Y.-A. *Nature* **360**, 757–759 (1992).
- Oppenheim, R. W., Qin-Wei, Y., Prevet, D. & Yan, Q. *Nature* **360**, 755–757 (1992).
- Schecterson, L. C. & Bothwell, M. *Neuron* **9**, 449 (1992).
- Ip, N. Y. *et al.* *Neuron* **10**, 137–149 (1993).
- Meyer, M., Matsuo, K., Wetmore, C., Olson, L. & Thoenen, H. *J. Cell Biol.* **119**, 45–54 (1992).
- Enfors, P. & Persson, H. *Eur. J. Neurosci.* **3**, 953 (1991).
- Wright, E. M., Vogel, K. S. & Davies, A. M. *Neuron* **9**, 139–150 (1992).
- Hughes, R. A., Sendtner, M., Goldfarb, M., Lindholm, D. & Thoenen, H. *Neuron* **10**, 369–377 (1993).
- Barres, B. A. *et al.* *Development* (in the press).
- Sendtner, M., Stockli, A. & Thoenen, H. *J. Cell Biol.* **118**, 139–148 (1992).
- Sendtner, M. *et al.* *Nature* **358**, 502–504 (1992).
- Leung, D. W. *et al.* *Neuron* **8**, 1045–1053 (1992).
- Martinou, J.-C., Martinou, I. & Kato, A. C. *Neuron* **8**, 737–744 (1992).

Another protein with CNTF-like activity, growth promoting activity (GPA), is found in chick sciatic nerves¹⁹; unlike CNTF, GPA is secreted but it is not yet known whether it acts on mammalian motor neurons *in vivo*. Fibroblast growth factor-5¹⁵, leukaemia-inhibitory factor^{1,20} and insulin-like growth factor⁴ also promote the survival of motor neurons *in vitro*, and seem to be made by muscle (see ref. 15), but they too have not been tested

for their ability to rescue motor neurons *in vivo*. Whatever their roles turn out to be in the life of motor neurons, there is now no shortage of proteins that can be tested for their therapeutic value in treating degenerating motor neurons. □

Anne W. Mudge is in the MRC Developmental Neurobiology Programme, Department of Biology, University College London, Gower Street, London WC1E 6BT, UK.

MAGNETIC RESONANCE

ESR on a single molecule

Robert J. Silbey

WITH condensed matter containing many billion billion molecules in every cubic centimetre, the newly acquired ability to pick out just one is remarkable. On pages 242 and 244 of this issue, respectively, J. Köhler *et al.* and J. Wrachtrup *et al.* describe experiments to obtain the electron magnetic resonance spectra of single molecules embedded in pentacene crystals. Because magnetic resonance spectra are sensitive measures of the shape of a molecule and of its interaction with surrounding molecules, this work opens up a new approach to molecular studies.

The experiments take advantage of the fact that in a solid at low temperature, the absorption spectra of individual guest molecules will have resonances at slightly different wavelengths owing to small differences in the intermolecular forces that each molecule experiences. This gives rise to a broadened spectral line made up of the sum of many single-molecule absorptions. In the wings of this broad line, the number of molecules contributing to the absorption decreases until far from the centre of the line, it is possible to measure (by fluorescence detection) the spectrum of a single molecule.

The authors of the new work have previously found that some molecules have a spectrum that remains constant in time, while others have a spectrum that changes (spectral diffusion). The latter effect was attributed to time-dependent fluctuations in the surroundings of the absorbing molecule. Thus, both the static energy shift due to the intermolecular forces and the time dependence of these forces can be studied by this technique. Now these groups have added an important new tool — fluorescence-detected electron-spin-resonance (ESR) spectroscopy.

The researchers again study the single-molecule spectrum by fluorescence detection, but this time they note that some of the molecules that are excited by light find themselves in a low-energy triplet state by the process of intersystem crossing. These

molecules will stay in the triplet state long enough for the three energy levels of the triplet state to be probed with microwave radiation. These three states have different lifetimes, so by changing the state of the triplet state using microwave radiation (in other words, performing an ESR experiment while the molecule is in the triplet state), the fluorescence intensity is changed. Thus the ESR transitions in a single molecule are detected through the optical fluorescence.

The line shapes of the ESR transitions are of interest because they are a measure of the interactions between nuclear and electron spins (hyperfine interactions). Because the hydrogen nuclear spins relax slowly when the molecule is in the triplet state, one might expect that the line shape would be a signature of a single nuclear spin state, and therefore much narrower than the normal ESR line. Unfortunately, to detect the spectrum with sufficient signal to noise, the experiment has to be repeated on the same molecule many times. When the molecule is in a singlet state (ground or excited), the nuclear spins relax faster by interacting with the other nuclear spins in the crystal. Thus the average of many experiments is an average over all the different nuclear spin states, giving rise to a line shape identical to that seen in the normal many-molecule experiment.

So the experiments at present offer the tantalizing prospect of improved kinds of magnetic resonance experiments, such as ENDOR and optically detected nuclear spin polarization, which will eventually allow the study of the effects of the environment on the electronic structure of single molecules in crystals and in disordered or amorphous materials. Furthermore magnetic resonance spectroscopy may be ideal for elucidating the factors contributing to spectral diffusion. □

Robert J. Silbey is in the Department of Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

DAEDALUS

A right mess

THE messiest substance that Daedalus knows is a certain two-part silicone adhesive. Insidiously tacky, and with a low surface tension, it sticks copiously to anything it touches. It manages to get onto the handle of any spoon or spatula used to mix or transport it. Daedalus attributes this to the characteristic ability of thixotropic liquids to climb a rotated rod.

This masterpiece of messiness has been formulated quite unintentionally, without even any concept of messiness as a property of viscous liquids. So Daedalus reckons that a deliberate search should turn up some staggeringly messy self-spreading fluids. A team of tough-minded DREADCO chemists is now on the quest. Fluids of incredible exasperation are emerging, or rather escaping, from their laboratory. Handrails and carpets, tools and instruments round the company are developing a disquieting tackiness, and the chemists themselves are being shunned as a clique of untouchables. Daedalus is clearly onto something important.

"Hypergunge" (as it will be called) will probably sell on its sheer novelty at first, as bouncing putty and borate slime did before it. But serious uses should soon appear, such as the tracing of criminal contacts. A transparent and inconspicuous Hypergunge would be loaded with a sensitive fluorescer, or a halocarbon detectable in minute traces by an electronic 'sniffer'. A thief or burglar who stole an object protected with Hypergunge would not realize that he was now covered with it. But the resulting trail could easily be followed by the police; it might even spread via the criminal to his contacts and associates.

More humanely, Hypergunge could be used to spread insecticides among a reluctant population. The more gregarious and unhygienic that population, and therefore the more at risk from infestation, the more relentlessly it will spread all over them and their property. The insects, too, would find it getting lethally all over them and their hiding places. Similarly, a Hypergunge sheep-dip applied to a few animals would treat the whole flock, not to mention the farmworkers, their possessions, and their friends and relatives. In psychiatry, Hypergunge would provide perhaps the ultimate challenge to the obsessional hand-washer. But novel containment and safety techniques will have to be developed. The consequences of a road accident to a tanker of Hypergunge do not bear thinking about. David Jones

Early medical use of cannabis

SIR — Records of the medicinal use of cannabis appear in the Egyptian Ebers papyrus of the sixteenth century BC. Much later, the plant is mentioned in Assyrian texts as a medicinal agent, and in Greek and Roman sources as a minor drug (see ref. 1). But physical evidence of cannabis use in the ancient Middle East has not yet been obtained. We found the skeletal remains of a girl aged about 14 at death in an undisturbed family burial tomb in Beit Shemesh, near Jerusalem. Three bronze coins found in the tomb dating to AD 315–392 indicate that the tomb was in use during the fourth century AD. We found the skeletal remains of a full-term (40-week) fetus in the pelvic area of the girl, who was lying on her back in an extended position, apparently in the last stages of pregnancy or giving birth at the time of her death.

The anterior–posterior internal dimensions of the immature pelvis of the young girl, which are of critical importance in the normal birth process, were approximately 7–7.5 cm. It is unlikely, though possible, that a normal vaginal delivery could take place when the anterior–posterior diameter of the pelvis is less than 9 cm. Thus it seems likely that the immature pelvic structure through which the full-term fetus was required to pass was the cause of death in this case, due to rupture of the cervix and eventual haemorrhage.

In the abdominal area of the skeleton, 6.97 g of a grey, carbonized material was recovered and analysed. The initial microscopic analysis, undertaken by the Israel Police Forensic Laboratory, revealed the possible presence of *Cannabis sativa*, and a further sample, examined by microscopy at the Department of Botany of the Hebrew University, showed the apparent presence of reed (*Phragmites*) and minute quantities of carbonized fruit or seed.

In our analysis, we separated 1 g of the grey powder and, by ether extraction, obtained a yellow oil (10 mg). Thin-layer chromatography (15% ether in petroleum ether, silica gel plates) revealed one main spot with an R_f value (0.72) identical with that of Δ^6 -tetrahydrocannabinol (Δ^6 -THC), a component of cannabis. We confirmed our conclusion using gas chromatography, mass spectrometry and NMR spectroscopy, all of which gave spectra

identical with those obtained from authentic Δ^6 -THC.

Δ^6 -THC is a minor, highly stable constituent of cannabis; its presence presumably indicates conversion of the major components, Δ^1 -THC and cannabidiol, into Δ^6 -THC, a process known to occur after acid catalysis, apparently during the burning process².

We assume that the ashes found in the tomb were cannabis, burned in a vessel and administered to the young girl as an inhalant to facilitate the birth process. In antiquity, this procedure would usually have been carried out by a midwife as physicians were by law prohibited from attending women in labour³.

As mentioned above, scattered literary references indicate that cannabis was used for a wide variety of medical conditions well into the twentieth century¹. In the nineteenth century, its use in ob-

stetrics is cited in several medical publications, one of which⁴ indicates that *C. sativa* had the remarkable power to increase the force of uterine contractions, concomitant with a significant reduction of labour pain.

Joe Zias

Harley Stark

Jon Seligman

*Israel Antiquities Authority,
PO Box 586, Jerusalem, Israel*

Rina Levy

Division of Identification and Forensic

*Sciences, National Police Headquarters,
Jerusalem, Israel*

Ella Werker

*Department of Botany, Hebrew University,
Jerusalem 91904, Israel*

Aviva Breuer

Raphael Mechoulam*

*Department of Natural Products,
Hebrew University Pharmacy School,
Jerusalem 91120, Israel*

* Author for correspondence.

World population forecasts

SIR — I read Tuckwell and Kozlols Scientific Correspondence¹ with a sense of *déjà vu*. The exercise of fitting the logistic curve to human populations was carried out for the US population on census data from 1790 to 1910 by Reed and Pearl and reported by Lotka² in 1925: "... if the population of the United States continues to follow this growth curve in future years, it will reach a maximum of some 197 million souls ... by the year 2060 or so". This "maximum" population was achieved by 1970 (ref. 3). Tuckwell and Kozlols extrapolation is based on only 35 years of census data but they predict a stable world population of 23.8 billion people by the year 2200. Two problems bedevil such curve fitting: one is that a short run of data will fit almost any curve; the other is the assumption that there is some fixed carrying capacity to which the human population is inexorably moving. The danger is that such predictions may be put to political use.

The limitation of the human population of the Earth is a pressing need, both for the reduction of human misery and for the wise management of the biosphere. The UN predictions⁴ criticized by Tuckwell and Kozlols recognize the basic equation of population dynamics:

$$\text{population growth} = \text{births} + \text{immigrants} - \text{deaths} - \text{emigrants}$$

for each age group and thus can predict not only population size but also age structure. Such estimates are normally useful for only 5–10 years, but they do lay bare the processes which should be studied.

The population of any organism is difficult to predict (except under close domestication) because it is subject not only to complex responses to a changing environment but also to evolutionary changes in phenotypic and genotypic composition. Human populations also respond rapidly to cultural, economic and technological change. Between 1966 and 1986, for example, the birth rate in China fell from 35.1 to 17.8 and the death rate from 8.8 to 6.6 (per 1,000) and the mean age of women at first marriage rose from 18.6 to about 23 years⁵. This was achieved through government-directed campaigns which promoted the one-child family, raised living standards, provided contraceptives and abortion, penalized large families and took more women into the workforce. In the United States a similar drop in birth rates, a rise in age at marriage and a corresponding change in social values occurred⁶.

Birth and death rates seem to be strongly dependent on the history of a country and have variable response times in the face of environmental change. However, if a global war is avoided, the following trends seem likely to stabilize world population after the current industrial revolution: (1) All countries are attempting to urbanize and industrialize their populations. If they do this in a framework of ecological stability the birth rates will ultimately fall. (2) Many countries are near to achieving maximum life expectancy. By the year 2050, population size will probably be regulated through changes in the number and timing of births.

1. Mechoulam, R. (ed.) in *Cannabinoids as Therapeutic Agents* 1–19 (CRC, Boca Raton, 1986).

2. Mechoulam, R. (ed.) in *Marijuana. Chemistry, Pharmacology, Metabolism and Clinical Effects* 1–99 (Academic, New York, 1973).

3. Myles, M. F. *Textbook for Midwives* 9th edn (Livingstone, London, 1981).

4. Christison, A. *Mon. J. med. Sci.* **13**, 26–45, 177–121 (1851).

There are dangers on the path to stability. Probably the most serious are delays in receiving, or distortion in communicating, environmental signals⁷. Ecologists have long recognized that the global stability of an animal population is increased by subdividing it into a set of self-sufficient local groups. For humans this would mean national self-sufficiency in essential foods. At the moment, imported food is stimulating population growth in some countries and thereby concealing and increasing a latent overshoot, which will one day be expressed. Despite such problems, the current crop of economists maintain the growth ethic and the notion that only price and the market should determine production and distribution. Perhaps it is time to subsume their discipline into ecology.

John M. Monro

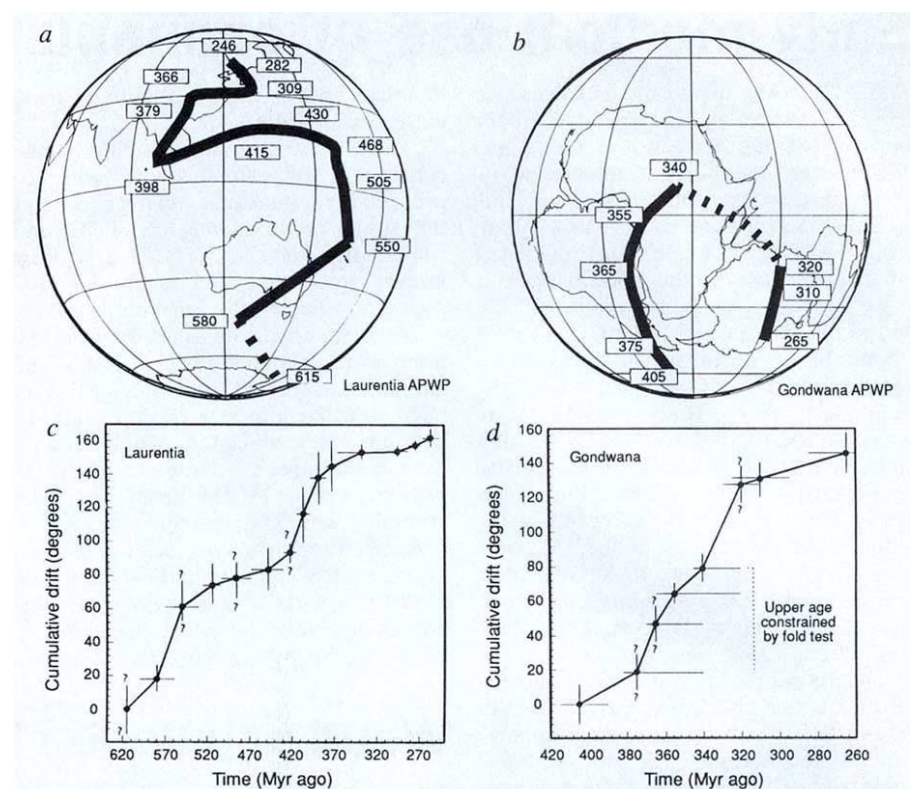
Department of Animal Science,
University of New England,
Armidale 2351, Australia

1. Tuckwell, H. C. & Kozlowski, J. A. *Nature* **359**, 200 (1992).
2. Lotka, A. *Elements of Physical Biology* (Williams & Wilkins, Baltimore, 1925).
3. Boughey, A. S. *Man and the Environment* (Macmillan, New York, 1971).
4. *World Population Prospects* (United Nations, New York, 1988).
5. Wang, J. & Hull, T. H. *Population and Development Planning in China* (Allen & Unwin, Sydney, 1991).
6. Weeks, J. R. *Population* (Wadsworth, Belmont, California, 1989).
7. Meadows, D. H., Meadows, D. L. & Randers, J. *Beyond the Limits* (Earthscan, London, 1992).

A plate-tectonic speed limit?

SIR — Estimates of the minimum velocities of lithospheric plates before the age of the oldest sea floor can be made using paths of apparent polar wander^{1,2}. Today, plates containing a significant amount of continental crust² ($> 2 \times 10^7$ km²) are known to move at speeds approaching those of oceanic plates (6–10 cm yr⁻¹). This contrasts with a suggestion by Forsyth and Uyeda³ that because of excess asthenospheric drag on continental plates, plate velocity should vary inversely with the amount of continental material.

A recent attempt⁴ to balance plate driving forces with observed plate velocities found that a best fit to the models occurs when 'continental drag' is eliminated; however, other models⁵ still call for a significant amount of asthenospheric drag on continental blocks. Klootwijk *et al.*⁶ inferred from sea-floor magnetic anomalies and palaeomagnetic data that India, as part of a mostly oceanic plate, moved at 19 cm yr⁻¹ from 80 to 50 million years (Myr) ago; more recently, Tarduno *et al.*⁷ and Gordon⁸ have suggested that the oceanic plate 'speed limit' could reach 30 cm yr⁻¹ — faster than speeds so far reported for



a, APWP for Laurentia (North Poles). Dates given are Myr ago. The path for Laurentia reflects new data from the Catocin formation (570 Myr, pole at 43° S 128° E), the Callander complex (577 Myr, pole at 44° S 134° E). The preliminary pole for the Manitou Islands complex (570 Myr) falls at 40° S 115° E supporting this portion of the path. The portion of the Laurentian APWP from 580–615 Myr is based on a single dyke dated at 615 Myr (pole at 69° S 170° E)¹³. This segment of the path is dashed to reflect the uncertainty. The 550 Myr portion is based on data from the Buckingham flows (pole at 16° S 165° E) and from the Long-Range dykes dated at 553 Myr (pole at 11° S 164° E). The remaining portion of the Laurentian APWP is based on path in ref. 14. **b**, APWP for Gondwana (South Poles) based on the analysis in refs 11, 12. The new Australian data have excellent fossil control with magnetic ages constrained by fold tests to within 60 Myr of deposition. **c**, Cumulative latitudinal drift curve for Laurentia. Drift rates can be calculated from the slope of the curve for any time segment. (Error bars indicate $\alpha 95$ about the mean poles and error in age; ?, error bar calculated from the $\alpha 95$ brackets for 1 or 2 poles only). **d**, Cumulative latitudinal drift curve for Gondwana.

large continental plates. We point out here that new palaeomagnetic data for Laurentia and Gondwana^{9–13} indicate that in the past large continental plates have travelled faster than 16 cm yr⁻¹, approaching the speed of small oceanic plates.

Previously published apparent polar wander paths (APWP) for Laurentia in the Cambrian period form a diffuse swathe prior to a well-constrained Late Cambrian (505 Myr old) pole⁹. New palaeomagnetic results from the Callander complex (Canada)¹⁰, Manitou Islands complex (Canada) (D. T. A. S., unpublished observations), and Catocin Formation (United States)⁹ were combined with earlier results from the Sept-Isles complex and Long-Range dykes in Canada. The new APWP indicates that Laurentia occupied polar latitudes during the interval 615–580 Myr ago followed by rapid drift to the equator during Vendian–Early Cambrian time (*a* in the figure). The figure (*c*) also shows the cumulative drift versus time curve for Laurentia 615–245 Myr ago. The

graph indicates two intervals of fast plate motion. Both intervals (580–550 and 415–379 Myr ago) give minimum peak rates of 16 cm yr⁻¹.

Chen *et al.*^{11,12} recently completed palaeomagnetic investigations of Late Devonian–Early Carboniferous rocks from the Amadeus and Ngalia basins (Australia). These results, combined with a re-evaluation of previous results for Gondwana 405–260 Myr ago lead to a revised APWP with the 375–340-Myr-old segment well represented by sequential poles (*b* in the figure). The 340–320-Myr-old portion of the path is constrained by a key pole 340 Myr ago and the mean of a group of poles 320 Myr ago. The figure (*d*) shows the cumulative drift curve for Gondwana 405–260 Myr ago, with drift rates in excess of 18 cm yr⁻¹ 375–355 and 340–320 Myr ago. Both the Laurentian and Gondwanan drift rates are minimum estimates because we have not included any longitudinal motion.

Do these palaeomagnetic data indicate that a fundamental change is needed in

our understanding of geodynamic forces that drive the plates? The suggestions^{3,5} that continental lithosphere has a greater resistance to motion and hence moves more slowly than oceanic lithosphere owing to asthenospheric drag are not supported by our data. The suggestion³ that the velocity of a plate is controlled by the length of subduction zones 'pulling' the plate cannot yet be tested, as we have no indication of the length of subduction zones bordering the Laurentia and Gondwana plates during their intervals of rapid motion.

We have yet to find an explanation for these periods of fast motion, but note that the duration of rapid drift for each of our examples does not exceed 50 Myr. There could have been intervals of faster motion, especially since we cannot determine longitudinal motion. Forsyth and Uyeda³ analysed only the most recent few million years of plate motions, perhaps giving rates biased by a period of mantle quiescence. Long-term fluctuations in mantle dynamics (for example superplume activity) could alter the balance of forces driving plate motions and allow continental plates to move at rates comparable to the fastest oceanic plates. Finally, true polar wander can always be invoked to explain the fast rates calculated for Laurentia and Gondwana. However, true polar wander is not significant for the past 80 Myr and, because there are currently no compelling data to indicate amounts of true polar wander for earlier times, we conclude that our rates closely approximate true minima.

Joseph G. Meert

Rob Van der Voo

*University of Michigan,
Ann Arbor, Michigan 48109, USA*

Chris McA. Powell

Zheng-Xiang Li

Michael W. McElhinny

Zhong Chen

*University of Western Australia,
Nedlands, Western Australia 6009*

D. T. A. Symons

*University of Windsor,
Windsor, Ontario, Canada N9B 3P4*

Fluctuating asymmetry

SIR — Sullivan *et al.* suggested in Scientific Correspondence¹ that the use of fluctuating asymmetry as an indicator of selective pressures on ornamental characters in animals is statistically flawed (see refs 2–5). The use of relative asymmetry, calculated as the absolute difference in the lengths of paired morphological characters divided by the mean length of the character, rather than absolute asymmetry dates back to 1986 (ref. 6). Relative asymmetry allows comparison of the asymmetry of the same trait in different species, such as the leg of a mouse and an elephant, or different traits in the same species, such as a leg and a tooth. Without controlling for the size of a trait, comparison between species will be uninformative; the elephant or the leg will always have larger asymmetry. Relativization should be with respect to the allometric relationship between characters, and that need not be linear. We used a linear relativization because the allometric relationship demonstrated isometry⁵. Regression of relative asymmetry on character size has only been used once⁴, and the conclusions of this analysis did not differ from those based on regression of absolute asymmetry on character size.

Sullivan *et al.*¹ suggest that a negative relationship between asymmetry and mean size of a character can arise simply because of consistent measurement errors. But this explanation does not apply to our previous results^{2–5}, because measurements of asymmetries were highly repeatable and measurement errors therefore relatively small. If Sullivan *et al.*¹ were correct, a spurious negative correlation between relative asymmetry and character size should arise more frequently in characters with relatively large measurement errors. However, the negative relationship between absolute asymmetry and mean character size is more often negative for secondary sexual characters with large asymmetries and relatively small measurement errors than for ordinary morphological characters with relatively large errors^{2–5}. This result invalidates the explanation of Sullivan *et al.*

Sullivan *et al.* also suggest that differences in character length may be due to unfinished growth, injury or damage, and that this effect may affect estimates of asymmetry. They suggest that the only rigorous approach is to use maximum rather than mean length of characters as the independent variable in statistical analyses. A more rigorous and straightforward approach is directly to

examine specimens for growth, injury or damage rather than basing the approach on untested assumptions. Exclusion of individuals with unfinished growth (birds with feather quills) or damaged characters (broken and worn feathers) has routinely been adopted in previous studies^{2,4,5}. The conclusions drawn from such studies are therefore less biased than would be the case if the approach of Sullivan *et al.*, with its untested assumptions about growth, damage and wear, had been adopted.

A. P. Møller

*Department of Zoology,
Uppsala University,
Box 561,
S-751 22 Uppsala, Sweden*

SIR — Sullivan *et al.*¹ rightly point to the dangers of statistical regression analysis where "the same measurement ... occurs as a component of both the dependent and independent variables, which can lead to spurious correlations". They claim that measurement error incorporated into the measurement in question can produce a correlation when none exists. They then go on to recommend a procedure where the absolute difference between two measurements is regressed on the larger of the two measurements. They have fallen into their own trap.

Paul H. Harvey

Sean Nee

Andrew F. Read

*AFRC Unit of Ecology and Behaviour,
Department of Zoology,
University of Oxford,
Oxford OX1 3PS, UK*

SIR — Fluctuating asymmetry describes the symmetrical distribution of random deviations from the population mean of a bilaterally symmetrical trait, presumed to arise from failure of homeostatic mechanisms in the face of developmental stress^{7–10}. As individuals vary in their ability to resist disruption of symmetrical growth, the relationship between trait asymmetry and trait length can give valuable insights into how selection acts on trait length¹⁰. Traits under stabilizing selection often show U-shaped relationships between asymmetry and mean trait length¹⁰, with extreme individuals poorly adapted to prevailing conditions. It is thus significant that recent studies indicate that many secondary sexual characters show negative relationships between asymmetry and size^{2,4,5,10–12}, indicating that sexual selection is directional and that individuals with large display traits are of 'high quality' (better adapted to prevailing conditions)¹⁰. Sullivan *et al.*¹ cast doubt on this evidence by suggesting two reasons for such negative correlations being artefacts. Their first criticism, of the use of relative asymmetry,

1. Ullrich, L. & Van der Voo, R. *Tectonophysics* **74**, 17–29 (1981).
2. Gordon, R. G., McWilliams, M. O. & Cox, A. J. *geophys. Res.* **84**, 5480–5486 (1979).
3. Forsyth, D. W. & Uyeda, S. *Geophys. J. R. astr. Soc.* **43**, 163–200 (1975).
4. Harper, J. F. *Geophys. J. Int.* **100**, 423–433 (1990).
5. Ricard, Y. & Wuming, B. *Geophys. J. Int.* **105**, 561–573 (1991).
6. Klootwijk, C. T. *et al. Geology* **20**, 395–398 (1992).
7. Tarduno, J. A., McWilliams, M. O. & Sleep, N. *J. geophys. Res.* **95**, 15503–15527 (1990).
8. Gordon, R. G. *Nature* **349**, 16–17 (1991).
9. Meert, J. G., Van der Voo, R. & Payne, T. J. *geophys. Res.* (submitted).
10. Symons, D. T. A. & Chiasson, A. D. *Can. J. Earth Sci.* **28**, 355–363 (1991).
11. Chen, Z. *et al. Geophys. J. Int.* (in the press).
12. Chen, Z., Li, Z.-X., Powell, C. McA. & Balme, B. E. *J. geophys. Res.* (submitted).
13. Murthy, G., Gower, C., Tubrett, M. & Patzgold, R. *Can. J. Earth Sci.* **24**, 1431–1438 (1992).
14. Van der Voo, R. *Rev. Geophys.* **28**, 167–206 (1990).

is well-founded, but we believe their subsequent suggestions to be misleading.

Relative asymmetry is the ratio of the absolute difference between the left and right side to the mean trait length. The problem arises here because relative asymmetry is necessarily negatively correlated with its denominator, mean trait length¹. Use of relative asymmetry will only 'control' for differences in mean trait length when absolute asymmetry is isometrically related to mean trait length. Under such circumstances, relative asymmetry may be a useful index, but isometry must be empirically verified, not assumed. However, belief in the negative correlation between asymmetry and mean trait length does not rest on the dubious use of a ratio variable¹³; Møller and his co-workers^{2,4,5,10,11} find the same relationship with absolute asymmetry, as do Sullivan *et al.* in their analysis of pheasant spurs¹.

The second criticism hints at a deeper problem in the analysis of asymmetry data, but the suggested solution is inappropriate and the true problem is not addressed. Sullivan *et al.* note that a shortening of one of a pair of bilaterally symmetrical characters "because of differences in growth rate, injury or damage" both increases absolute asymmetry and reduces mean length, thereby biasing the analysis toward a negative correlation between the two. Their solution is to examine the relationship between absolute asymmetry and the longest of the pair of traits, rather than the mean length. However, this reasoning is flawed on two counts. First, unlike asymmetry through damage or wear, true fluctuating asymmetry (imbalances in symmetrical growth) can result from overgrowth on one side as well as stunting on the other. Thus higher asymmetries through a failure of developmental homeostasis do not inevitably lead to a reduction in the mean trait length, merely an imbalance between the two sides. Second, analysis of the relationship between absolute asymmetry and the longest of the two traits in fact biases one towards finding a positive correlation, as we show below.

The relationship between absolute asymmetry and trait length can be examined in a number of ways. Sullivan *et al.* consider two: "asymmetry versus mean" and "asymmetry versus longest". However, not all such comparisons are equivalent, or appropriate. In the language of principal component analysis, the mean trait length represents the dimension in which the left and right sides are correlated and absolute asymmetry is the orthogonal dimension in which they are not. However, the dimensions "longest side" and "absolute asymmetry" are not orthogonal,

but positively correlated, thus biasing the analysis. Intuitively, this is most easily understood in the situation where left and right sides are uncorrelated. When the longest side is much greater than the population mean, the short side can assume almost any value (hence asymmetry can be large); but when the longest side approaches the lower limit of the trait size distribution, the shorter side is more constrained (hence asymmetry is low). The problem is greatest when the correlation of left and right sides is relatively loose; precisely the situation that appears typical of display traits^{4,5,10}.

In recognizing that damage and wear (but not developmental stress) lead to both an increase in asymmetry and decrease in mean length, Sullivan *et al.* raise an important point that deserves amplification. As only fluctuating asymmetries, resulting from developmental stress, are of relevance to investigations of the direction and strength of selection, it is vital to exclude any measurements influenced by wear or damage. Sullivan *et al.* may not be able to distinguish true fluctuating asymmetries from damage or wear in their pheasant spurs, but with sexual plumage, breakages and wear are probably easier to identify *a priori*. With mean asymmetry

more sensitive to such outliers than mean trait length, and ornaments perhaps particularly prone to damage, it is all the more important to exclude non-developmental asymmetries from analyses of fluctuating asymmetry. If one can exclude damage and wear effects, an analysis of the relationship between absolute asymmetry and mean trait size is both entirely appropriate and of great theoretical importance¹⁰.

Innes C. Cuthill

John P. Swaddle

Mark S. Witter

*School of Biological Sciences,
University of Bristol,
Bristol BS8 1UG, UK*

1. Sullivan, M. S., Robertson, P. A. & Aebischer, N. *Nature* **361**, 409–410 (1993).
2. Møller, A. P. *Anim. Behav.* **40**, 1185–1187 (1990).
3. Møller, A. P. *Nature* **357**, 238–240 (1992).
4. Møller, A. P. *Proc. R. Soc. B* **248**, 199–206 (1992).
5. Møller, A. P. & Höglund, J. *Proc. R. Soc. B* **245**, 1–5 (1991).
6. Palmer, A. R. & Strobeck, C. A. *Rev. Ecol. Syst.* **17**, 391–421 (1986).
7. Ludwig, W. *Das Rechts-Links Problem im Tierreich und beim Menschen* (Springer, Berlin, 1932).
8. Van Valen, L. *Evolution* **16**, 125–142 (1962).
9. Parsons, P. A. *Biol. Rev.* **65**, 131–145 (1990).
10. Møller, A. P. & Pomiankowski, A. *Genetica* (in the press).
11. Møller, A. P. *Proc. R. Soc. B* **243**, 59–62 (1991).
12. Manning, J. T. & Hartley, M. A. *Anim. Behav.* **42**, 1020–1021 (1991).
13. Raubenheimer, D. & Simpson, S. J. *Entomol. exp. appl.* **62**, 221–231 (1992).

Predicting Earth's lifespan

SIR — Caldeira and Kasting¹ suggest that Earth's biosphere will exist for more than 1,500 million years, instead of 100 million years as previously suggested². We would like to point out another alternative not considered by these authors.

Caldeira and Kasting did not take into account the CO₂ oscillations of recent Earth history^{3,4}. The causes of glacial-interglacial oscillations of about 100 p.p.m. are not clear, but they show the possibility of abrupt (on geological timescales) CO₂ changes of considerable amplitude. If such CO₂ oscillations are a natural behaviour of the global carbon cycle, with the lower mean atmospheric CO₂ contents expected in the distant future, very cold climates could ensue. These cold climates might also result if CO₂ degassing diminishes with lower geothermal fluxes and slower mantle mixing — an expected consequence of diminishing radioactive decay rates in Earth's mantle⁵.

These factors might lead to a much shorter lifespan for the biosphere because of catastrophic global glaciation, leading to the so-called 'white Earth' solution of energy-balance models^{6,7}. The climate of the 'white Earth' can be characterized by very low mean surface temperature (190–240 K) and a thick

(1,500–3,000 m) layer of ice in the global ocean⁸. This ice sheet would probably not be stable⁹, but such catastrophic glaciation would probably eliminate all higher forms of life. Geothermal and hydrothermal activity would probably continue to support chemotrophic and heterotrophic communities under permanent ice cover.

In conclusion, as far as the 'megabiosphere' is concerned, we concur with Robert Frost's conclusion in 1923 that whereas "some say the world may end in fire, some say ice . . . that for destruction ice is also great and would suffice".

Andrei Lopenis

Michael R. Rampino

*Earth Systems Group,
Department of Applied Science,
New York University,
New York, New York 10003, USA*

1. Caldeira, K. & Kasting, J. *Nature* **360**, 721–723 (1992).
2. Lovelock, J. E. & Whitfield, M. *Nature* **296**, 561–563 (1982).
3. Lorius, C. *et al.* *Nature* **347**, 139–145 (1990).
4. Chappellaz, J. *et al.* *Nature* **345**, 127–131 (1990).
5. Hubbard, W. B. in *Planetary Interiors* (Van Nostrand Reinhold, New York, 1984).
6. Budyko, M. I. *Tellus* **21**, 611–619 (1969).
7. North, G. R. *J. Atmos. Sci.* **41**, 3390–3395 (1975).
8. Verbitsky, M. Ya. & Chalikov, D. V. *Modeling of the Glaciers — Ocean-Atmosphere System* (ed. Monin, A. S.) 132 (1986) (in Russian).
9. Caldeira, K. & Kasting, J. *Nature* **359**, 226–228 (1992).

Prisoners of fate

Brian Bertram

The Last Panda. By George B. Schaller. *University of Chicago Press: 1993. Pp. 291. \$24.95, £19.95.*

WILL living giant pandas grace any part of our planet in our grandchildren's time? Not unless things are managed very differently from the shambles of recent years. And what a reflection that deplorable state of affairs would be on mankind — the loss of this famous symbol of nature conservation.

Since 1980, George Schaller has been more closely involved in panda conservation than almost anyone. He carried out the first significant field study of giant pandas in the early 1980s, the main research findings of which were published in *The Giant Pandas of Wolong* in 1985. Now, in *The Last Panda*, he gives the background, detail and realities of the crucial nonscientific issues. After further panda surveys, upheavals in China, chaotic developments in the captive panda world and a great deal of thought and heart-searching, Schaller lays out his recollections, emotions, wisdom and fears about pandas and their conservation. And he writes superbly — the book is quotably lively, varied and jumbled, riveting and often depressing, but essential reading to find out what the real world out there is like.

The fieldwork was (relatively) straightforward — “easy in conception and execution”, Schaller says — despite the terrain, the climate and the elusiveness of his subjects. These overgrown carnivores-turned-vegetarians exploit the bamboo niche, a food supply that is abundant and of constant but low nutritional value. With the limited gut capacity of a carnivore, a panda can use only the cell contents of the bamboo, not the cellulose of the cell wall as ruminants can. A panda therefore eats 25–30 pounds of bamboo a day, from 3,000–4,000 stems, spending 13 hours doing so and having little spare time or energy to do much of great interest. Nonetheless, there is a magical fascination about the animal and its habitat.

Pandas face an array of conservation problems. As everywhere, habitat destruction by humans is one of the greatest. “Sichuan's forests, the second largest storehouse of trees in the country, are rapidly disappearing. Such a loss of forests means death to the panda.” Fragmentation of the panda population into small isolated groups could eventually lead to genetic problems, but Schaller writes: “If what I had seen in Wolong and Jiuzhaigou was indicative, then poachers would eliminate the panda long before that”. Poaching in protected forests is rife,

and law enforcement weak. Smuggled panda pelts attract enormous prices, so “the killing continues, subsidized by the wealth of Hong Kong, Taiwan and Japan”. And the high public profile of pandas meant that they received excessive attention and often unnecessary ‘rescuing’ into inadequate captive conditions when

of the Cultural Revolution. Schaller writes: “Everyone was his neighbour's policeman. No one dared visit our tent alone, no one wanted to seem overly friendly to a foreigner; fear of denunciation pervaded the camp”. And, with a few fine exceptions, the Chinese fieldworkers allocated to the study were understandably uninterested; worse, two of them almost wrecked it with “their crafty coalition of misapplied intelligence and small-minded conspiracies”.

So what of the future? “With extraordinary tolerance for the burdens of disillusion, the WWF still continues its commitment to the panda.” At least there is now a conservation strategy for pandas,



Staff from the Wolong Reserve with a panda killed in a poacher's snare.

the periodic flowering and die-off of some bamboo species locally threatened their food supply.

The whole panda conservation programme in China was beset with bureaucratic and cultural problems. The World Wide Fund for Nature (WWF), the primary source of funds for the original programme, seemed ill-prepared for the negotiations and nuances involved in dealing with Chinese officials. The price extracted from a reluctant WWF for the acceptance of a much-needed field study and survey was the funding of an elaborate and unnecessary research centre, wanted by virtually no-one and little used since its construction; that, and the arguments and misunderstandings surrounding the centre, hung like an albatross round the project's neck.

Misunderstandings were easy enough anyway, even without language barriers and cultural differences. China was only just emerging from the ten years of chaos

although its details and status are unclear. We all have to hope that it will be implemented. To do so, China must get its act together; not easy, with local people ignoring distant government instructions. So must the West, where competition and arguments, and even lawsuits among conservation bodies over filming, publicity and rent-a-panda deals, have muddled the issues. Schaller observes with irony that “never has the panda's destruction been as rapid as during the years we studied it, during a period when it received more attention than at any time in its long history”. But persistent, patient effort, with open, informed and understanding eyes, not despair, is what is needed. It is what Schaller has contributed and is what the remarkable panda deserves. □

Brian Bertram was formerly curator of mammals at the Zoological Society of London and director-general of the Wildfowl and Wetlands Trust.

But today, there is naming of parts

J. S. Jones

Keywords in Evolutionary Biology.

Edited by Evelyn Fox Keller and Elizabeth A. Lloyd. *Harvard University Press*: 1992. Pp. 414. \$45, £39.95.

THE Maoris, it is said, have 35 different words for dung. To Maori philologists, if such exist, each term has — no doubt — an exquisitely distinct meaning. To the rest of us, dung is, well, dung. We may occasionally use different words to describe it, but it's all the same stuff really.

If you thought that the same was true for familiar biological stuff such as EXTINCTIONS, GENES or INDIVIDUALS, this book shows how wrong you were. Each — as do most of the 40 keywords discussed here — turns out to have hidden assumptions and ambiguities; so much so that some may be best abandoned.

Any science, such as evolution, that is still trying to define its terms more than a century after it began has real problems. Part of the semantic quandary arises from evolutionists' habit of picking up words that have already been used by someone else. COMPETITION, ADAPTATION, STRUGGLE and ALTRUISM: we all know what they signify in daily life, and easily slip into thinking that they must mean the same in science. The vaguer the concept, the easier it is to use. Muddled thinking has a distinguished pedigree. After all, Darwin himself admitted to using "the struggle for life" in a "large and metaphorical sense".

In evolution, there has always been a tendency for words to take precedence over thought. Take Herbert Spencer. In his obituary, the London *Times*, with the shoddy judgement which has characterized that newspaper ever since, wrote that "England has lost the most widely celebrated and influential of her sons . . . breathing life into the dead bones of science". Spencer was, without doubt, the most famous Darwinist of his day (although, according to P. Bowler's definition in *Keywords*, he was more a LAMARCKIAN, as he believed that the struggle for existence caused individuals to improve themselves and that this improvement was passed to generations yet unborn).

Spencer's definition of EVOLUTION was "an integration of matter and a concomitant dissipation of motion; during which the matter passes from an indefinite, incoherent homogeneity, to a definite, coherent heterogeneity, through continuous differentiations and integrations". The Victorian mathematician Kirkman parodied this (and it deserves

parody) as "a change from a nohowish untalkaboutable all-alikeness to a somehowish and in general talkaboutable not-all-alikeness by continuous something-elsefications and sticktogetherations".

But what *is* evolution? In *Keywords*, R. Richards points out that its meaning has gone full circle since it began. In its Latin root it referred (as readers of *Nature* will of course know) to the unfolding of a scroll. As Cicero so memorably put it: "*Quid poetarum evolutio voluptatis affert*" — What pleasure does the reading of the poets provide! Swammerdam used the word first in biology to refer to the unfolding shape of an insect during its metamorphosis, in his view the expression of a fixed adult form which was preformed in the egg. In the eighteenth century, as the idea of preformation declined, it shifted to mean the expression in a developing embryo of the structures of more primitive species — what is now referred to as recapitulation. Then it changed again to mean something close to the way in which it is now used, the gradual transformation of one species into another.

SPECIES, it turns out, is so hard to define that it needs three separate and conflicting interpretations. P. F. Stevens makes a historical survey, concentrating on the taxonomists' view of a species as just a rank in the great chain of being. J. Dupré asks whether members of a species belong to a natural set — or are they a kind of individual? He does not come up with an answer (although he manages to take a swipe at the biological species concept on the way). M. B. Williams takes a more relaxed view of how the word is used today, but also fails to find the key. Some modern definitions have an oddly Spencerian ring — one, from 1989, has it that a species is "the most inclusive population of individuals having the potential for phenotypic cohesion through intrinsic cohesion mechanisms", and there is, with Spencer, the same sense of trying to define the undefinable (which would mean, in any other science, the unusable). All this lexical confusion makes it easy to forgive Darwin for having written a book about almost everything except the origin of species.

Well, what about EXTINCTION? Simple enough, surely — deceased, no more, an ex-species. But no; as J. Damuth points out, there are subtleties even in being dead. The mere fact of extinction was once hard for creationists to bear, and President Thomas Jefferson enjoined the explorers Lewis and Clark to keep a keen look-out for the mammoths that must be roaming the United States somewhere. In spite of the 4,000-year-old stragglers reported a few weeks ago in *Nature*, the mammoths do seem to have departed. Nevertheless, even for a

living species, extinction is the norm, as most of its members are defunct. In addition, the whole idea of evolution assumes that even when a species has disappeared for ever, some of its descendants are flourishing. Such species are, apparently — and rather like Monty Python's parrot — only PSEUDOEXTINCT.

When we get to GENE it is beginning to be time to give up hope. P. Kitcher's definition is that a gene is anything a competent biologist chooses to call one. At least this definition is simple — unlike MONOPHYLY: "T is monophyletic, if and only if there exists a species s, a member of T; and for any other species d, if d is a member of T, then s is an ancestor of d".

A lot of this might seem like nit-picking, but most of it is essential. It is difficult to argue about speciation or group selection without everyone being sure that they are talking about the same thing. *Keywords* makes it clear that for much of the time they are not. Much of the problem arises from the fact that many terms — ADAPTATION, COMPETITION and NICHE, for example — include both process and pattern: that which exists, together with how it was attained. Sometimes what seems a trivial point of definition alters the entire argument. A failure to understand the precise meaning of HERITABILITY led to the whole wearisome controversy about nature and nurture in controlling human intelligence.

At an American sociologists' congress a few years ago, 'love' (not mentioned here) was defined as "the cognitive affective state characterized by intrusive and obsessive fantasising concerning reciprocity of amorous feelings by the object of amorance". It's better than saying that it is like a violin, but it's not much help to those faced with the problem. The same might be said of this book — it helped me to understand why I feel so confused about evolution, but did not come up with what the right answers might be. □

J. S. Jones is in the Department of Genetics and Biometry, University College London, London NW1 2HE, UK.

New Journals issue

This year, *Nature's* annual new journals review supplement will appear in the issue of 7 October. Publishers and learned societies are invited to submit journals for review. Journals that first appeared during or after June 1991 and issued at least four separate numbers by the end of April 1993 will be considered. Frequency of publication must be at least three times a year. For further information please see the 22 April issue of *Nature*, or telephone Peter Tallack on 071-836-6633 (011-44-71-836-6633 from the United States), extension 2414.

Circular arguments

H. Hopf

The Kekulé Riddle: A Challenge to Chemists and Psychologists. Edited by John H. Wotiz. Cache River Press, Rt. 3, Box 239c, Vienna, Illinois 62995, USA: 1993. Pp. 329. \$78.

ON 11 March 1890 a sumptuous celebration, later to be known as the *Benzolfest*, took place in Berlin City Hall to commemorate the twenty-fifth anniversary of the establishment of the benzene formula by Friedrich August Kekulé in 1865. A hundred years later, the American Chemical Society in turn held a symposium to commemorate the *Benzolfest* and in particular to discuss the scientific, political and psychological aspects of Kekulé and the early days of aromatic chemistry. *The Kekulé Riddle* assembles the text of most of the lectures given at this conference together with a few additional articles.

Kekulé (1829–96) is one of the few chemists known not only to scientists but also to the general public. The benzene hexagon is probably the most widely accepted symbol in chemistry, enhanced by the exciting story of its ‘formulation’ by Kekulé in a dream, not by logical reasoning. The book’s 19 chapters may be roughly grouped around five themes. First are the purely historical contributions about the *Benzolfest* and its organizers, the institute of chemistry that Kekulé established at the University of Bonn and the Belgian scientific community. Although they help to illustrate the social and scientific life of the late nineteenth century, these sections will appeal mainly to the professional historian of chemistry rather than the general reader.

From here the book moves to the beginnings of spectroscopy, and the early ideas about aromaticity and aromatic substitution. Modern chemists will probably not grasp the true difficulty of the problem faced by the early workers, who were unable to solve the relationship between structure and reactivity. A case in point is the Armstrong benzene formula, which — although incorrect — can account for the products formed in an electrophilic aromatic sub-

stitution. Not until the advent of quantum chemistry could this question be properly settled.

A third section concentrates on the scientific and personal relationships between Kekulé and his competitors in the ‘race’ to find the correct benzene structure, the main players being H. E. Armstrong, A. M. Butlerov and E. Frankland. This is followed by a pair of thought-provoking ‘psychological’ chapters on problem solving and creativity in general, written by a clinical psychologist and a psychiatrist respectively.

But the most provocative part of the book consists of two ‘revisionist’ chapters that set out to ‘demask’ Kekulé. Christian Noe and Alfred Beder argue that the Austrian chemist Johann Joseph

as proof of a “selective memory” rather than an example of normal human oversight. (Ironically, the authors themselves provide an incorrect reference to the volume of *Berichte Deutschen chemischen Gesellschaft* in which the *Benzolfest* is described — it is volume 23, not 22.) More important are the accusations that put Kekulé in the same political camp as the chemist Hermann Kolbe, whose strong nationalism and anti-semitism is well known. Kekulé, who spoke several languages and studied and taught in France, England and Belgium, had in fact been accused of “internationalism” by Kolbe and others. To depict him as a warmongering Prussian is ridiculous.

Signatures of Kekulé from between 1860 and 1882 are presented as proof that he tried to hide the French *accent aigu* on his surname as the antagonism between Germany and France increased. As far as I can see, the accent is more pronounced in the 1882 signature than in the very early one. Besides, signatures do change with age.

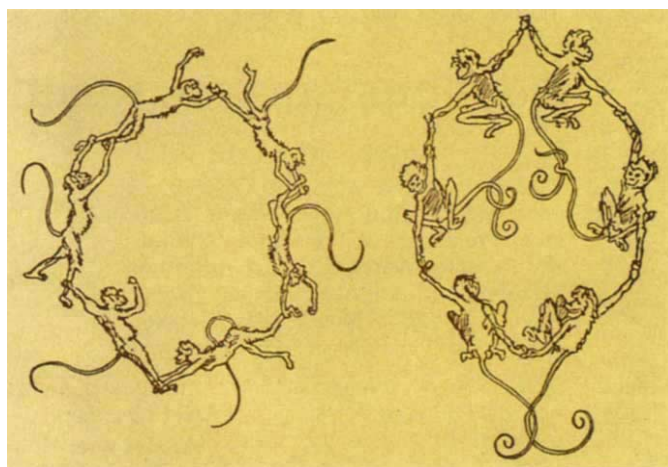
In all these rather strained ‘proofs’ one recognizes the intention of reaching a predetermined goal — to question Kekulé’s integrity.

Did Kekulé use the observations and concepts of others? Of course he did — after all, he had the best understanding of aromatic chemistry in his time. How many truly original ideas are there? Cannot one trace early forms of the electronic theory of organic structure and reactivity to Thiele’s ‘partial valencies’?

Are not some of the conformational drawings by Sachse and Mohr of the same (or even better) quality than those formulae used more than 50 years later? Of course they are. The list of early, but not really understood, discoveries of important phenomena is nearly endless. Indeed, Kekulé was aware of this when he used the ‘giants aphorism’ in his *Benzolfest* speech: “We all stand on the shoulders of our predecessors; is it then surprising that we can see further than they?”

The riddle detected by the editor of this book is a mere artefact. Kekulé is an enigma only in the sense that every human is. □

H. Hopf is in the Institute for Organic Chemistry, Technical University of Braunschweig, Hagenring 30, D-3300 Braunschweig, Germany.



Benzene structure according to F. W. Findig (a pseudonym of O. N. Witt). The parody, demonstrating how Kekulé’s formula can be derived on a zoological basis, was distributed at a meeting of the German Chemical Society in 1886. The carbon atoms are symbolized by monkeys whose tails, alternately free and clasped, reproduce the behaviour of the double bonds. Taken from *Reflections on Symmetry in Chemistry* by E. Hellbrunner and J. D. Dunitz (VCH, DM58, £22; see *Nature* 362, 670; 15 April 1993).

Loschmidt should be given the credit for solving the structure of benzene. Although Loschmidt drew circular formulae that show some resemblance to the geometrical structures used to depict the molecules today — and among them the so-called formula 185 for benzene — he never used the hexagon symbol. Readers interested in the subtleties of the Loschmidt formulae are referred to a recent paper by G. P. Schiemenz, himself a knowledgeable contributor to this book (G. P. Schiemenz, *Naturwissenschaftliche Rundschau* 46, 85–88; 1993).

The chapter by J. Wotiz and S. Rudofsky is supposed to be a documentation of “all of Kekulé’s misdeeds”, showing that he is “guilty of misconduct in science”. These are strong words, which in my opinion are hardly supported by the facts. For example, that Kekulé occasionally forgot references to other people’s publications is taken

Nature's way

Antonia J. Jones

Genetic Programming: On the Programming of Computers by Means of Natural Selection. By John R. Koza. MIT Press: 1992. Pp. 819. \$55, £49.50.

In 1975, John Holland published his classic book on adaptation in natural and artificial systems. His idea was that evolution by means of natural selection can be simulated on a computer and used to optimize the design of any given structure. His work differed from earlier attempts in two important respects. First, he capitalized on the idea that sexual reproduction is a critical accelerator of the evolutionary process and built this feature into his genetic algorithms using a 'crossover' operator that mixes the genes of both parents. Second, he gave genetic algorithms a theoretical credibility by showing mathematically why they could be expected to work better than, say, a random search.

It was inevitable that these ideas would eventually be applied directly to breed programs themselves, and Koza's book is the first substantial step in this direction. Holland's theory does not apply directly to Koza's scheme and Koza does not attempt to erect a theoretical edifice; he seeks only to demonstrate that the method does, in some sense, work. Major software systems are the most complex structures we have ever sought to engineer, and their reliability is often crucial. Consequently, formal specification and proof of correctness are important. The idea of a serious scientific study devoted to producing formally *incorrect* programs may therefore seem like heresy. Nonetheless, that is exactly what Koza and his students are doing and the results are rather amazing.

Obviously, if one simply splices bits of code together from two programs the chances that the resulting structure will even run, let alone do anything useful, are vanishingly small. Koza overcomes this difficulty by working in LISP, in which the program is essentially its own parse tree (a structure that demonstrates how a given, syntactically correct sentence may be derived from the rules of the underlying grammar). So a crossover that grafts a subtree from one program into another will at least produce a syntactically correct child. Of course, this is a convenience and not a restriction of generality. Having created a child program using crossover, the result is then run, once or many times, so that its fitness (as defined by criteria laid down at the start) can be automatically judged. The fitness is used to calculate the probability that the program will be a parent

in the next generation.

The criterion of fitness is analogous to a specification, and Koza chooses example problems for which this criterion is relatively easy to write down. Some are classic test problems in artificial intelligence, such as cart centring, broom balancing, truck reversing and box stacking. Others involve a pursuer-evader game, the control of a two-link robot arm, emergent collective behaviour of ants, task prioritizing (how to play Pac-Man) and the spontaneous emergence (in a nontrivial way) of self-replicating and self-improving programs. The 'Genetic Programming Paradigm' solves them all successfully, in the sense that workable, reliable programs emerge fairly rapidly.

To argue that it is hard to imagine the method being used to produce (say) a

spreadsheet is to miss the point. Producing software in cases where the rules that map input to output are precisely known is much less of a problem than in cases where the circumstances faced by the software are hard to foresee and the rules for dealing with these circumstances are largely unknown. Many of the central problems in artificial intelligence fall into this category and Koza's scheme offers an intriguing new possibility for automatic programming. The book is a *tour de force* of carefully presented case studies and includes LISP code with which the doubtful can verify the author's results. □

Antonia J. Jones is in the Department of Computing, Imperial College of Science, Technology and Medicine, Exhibition Road, London SW7 2BT, UK.

Aqueous surface reactions

William H. Casey

Chemistry of the Solid–Water Interface: Processes at the Mineral–Water and Particle–Water Interface in Natural Systems. By Werner Stumm. Wiley: 1992. Pp. 428. £34.50, \$48.95 (pbk).

AQUEOUS geochemistry is undergoing an upheaval forced, in part, by increasing pressure on governments to assess the safety of waste repositories and the fate of agricultural chemicals. For many years, processes taking less than 10^3 years or so were considered breathlessly fast. Scientific emphasis was placed on thermodynamic calculations that elucidated possible, rather than actual, pathways in the chemical history of the Earth. The length and timescales were simply too vast for close testing of the predictions.

The focus is now changed. Predictions about contaminant migration in the Earth are regularly tested but generally fail. The mathematics is daunting because sets of nonlinear equations describing chemical equilibria must be solved simultaneously with the equations of groundwater flow. Emphasis has been placed on *ad hoc* simplifications that reduce the mathematics to a linear form (but the chemistry to near absurdity). These simplifications work best when they describe mineral–fluid surface reactions, which are usually rapid and extensive.

Geochemists have established a nascent cottage industry adapting Werner Stumm and James Morgan's *Aquatic Chemistry* (Wiley, 1981) for use as a guide to these low-temperature processes. This is a wonderful monograph with a page or two devoted to

each aspect of natural water chemistry. But it is comprehensive to the point of being monolithic (780 pages): undergraduate students take one look and scurry off to study materials science.

In *Chemistry of the Solid–Water Interface*, Stumm expands the discussion of aqueous surface chemistry. He takes the familiar approach of treating surfaces as assemblages of coordinative functional groups. The chemistry of aqueous–metal complexes provides a key to understanding reactions at these functional groups.

The real strength of the new book, as in *Aquatic Chemistry*, is its breadth. The author goes well beyond simple reactions at oxide mineral surfaces to embrace hydrophobicity, polyelectrolytes and simple models of contaminant migration. Even bacterial adhesion is covered. Such discussion is timely: the attention of geochemists is increasingly drawn to environments that are contaminated with solvents and halogenated hydrocarbons and away from purely inorganic systems.

The book is, however, more a review of the author's interests and collaborations than an integrated text. This narrow focus is perhaps forgivable, as the book marks the culmination of Stumm's fruitful career as director of the Swiss Federal Institute of Technology. One need only ponder Carrick Eggleston's magnificent cover image of atomic steps on goethite (an iron oxide mineral) *in situ* to realize that extraordinary advances punctuate our progress in understanding natural water chemistry. □

William H. Casey is in the Department of Land, Air and Water Resources, University of California, Davis, California 95616-8627, USA.

Primate origins: plugging the gaps

Robert D. Martin

Recent discoveries of fossil primate specimens have produced several surprises and challenged prevailing views of early primate evolution. Plesiadapiforms, long regarded as 'archaic primates', may perhaps be linked to the peculiar colugos instead. Inferred relationships of the earliest known undoubted primates (adapids and omomyids) are in turmoil. Both groups have been proposed as sources for the simian primates. Although the origin of the simian primates is obscure, new fossil evidence could push it further back by at least 10 million years. Such uncertainties reflect the low sampling level of the primate fossil record, which can potentially also lead to underestimation of times of origin within the primate tree.

A PICTURE of the earliest phases of primate evolution that seemed broadly acceptable just a decade ago has been increasingly challenged. Several recent discoveries of early fossil primates have opened up new possibilities for the pattern and timing of emergence of the main groups of primates between the late Cretaceous and the end of the Eocene. As part of the upheaval, the cohesion of various fossil groups as monophyletic units has also been questioned. Widespread acceptance of the essentially Palaeocene Plesiadapiformes ('archaic primates') as a well defined group of exclusive early relatives of primates has been shaken by new fossil evidence and analyses. Certain features have been interpreted as indications that some genera, at least, may instead be directly linked to the Dermoptera (gliding colugos, also called 'flying lemurs'). New fossil specimens, especially postcranial parts, of early Tertiary Adapidae ('lemuroids') have increasingly confirmed that some European genera (*Adapis*, *Leptadapis* and relatives) belong to a distinct sideline relative to other European genera (for example *Europolemur*, *Pronycticebus*), which are morphologically closer to North American adapids (for example *Notharctus*, *Smilodectes*). A revived suggestion that simian primates (monkeys, apes and humans) may be derived from the Adapidae directly conflicts with an alternative interpretation that the early Tertiary Omomyidae ('tarsioids') are relatives of tarsiers and simians whereas adapids are specific allies of modern lemurs and lorises (strepsirrhine primates). Yet the discovery of several skulls of the North American omomyid *Shoshonius* led to the opposing claim that this genus is in fact even more directly linked to tarsiers, at the same time indicating a much earlier date for tarsier origins. Finally, dental evidence of the new genus *Algeripithecus* from a North African site dating back to the middle or even early Eocene would increase the age of the earliest yet identified simians by a striking margin of at least 10 million years. Given all this new fossil material and the sometimes conflicting interpretations that have been made, can the outlines of an emerging revised view of early primate evolution be traced? Consideration of the new evidence in the light of exploratory estimates of the current sampling level of the primate fossil record (Box 1) suggests that we may still have much to learn about the early relationships of primates and that inferred times of origin are likely to be pushed back much further into the past.

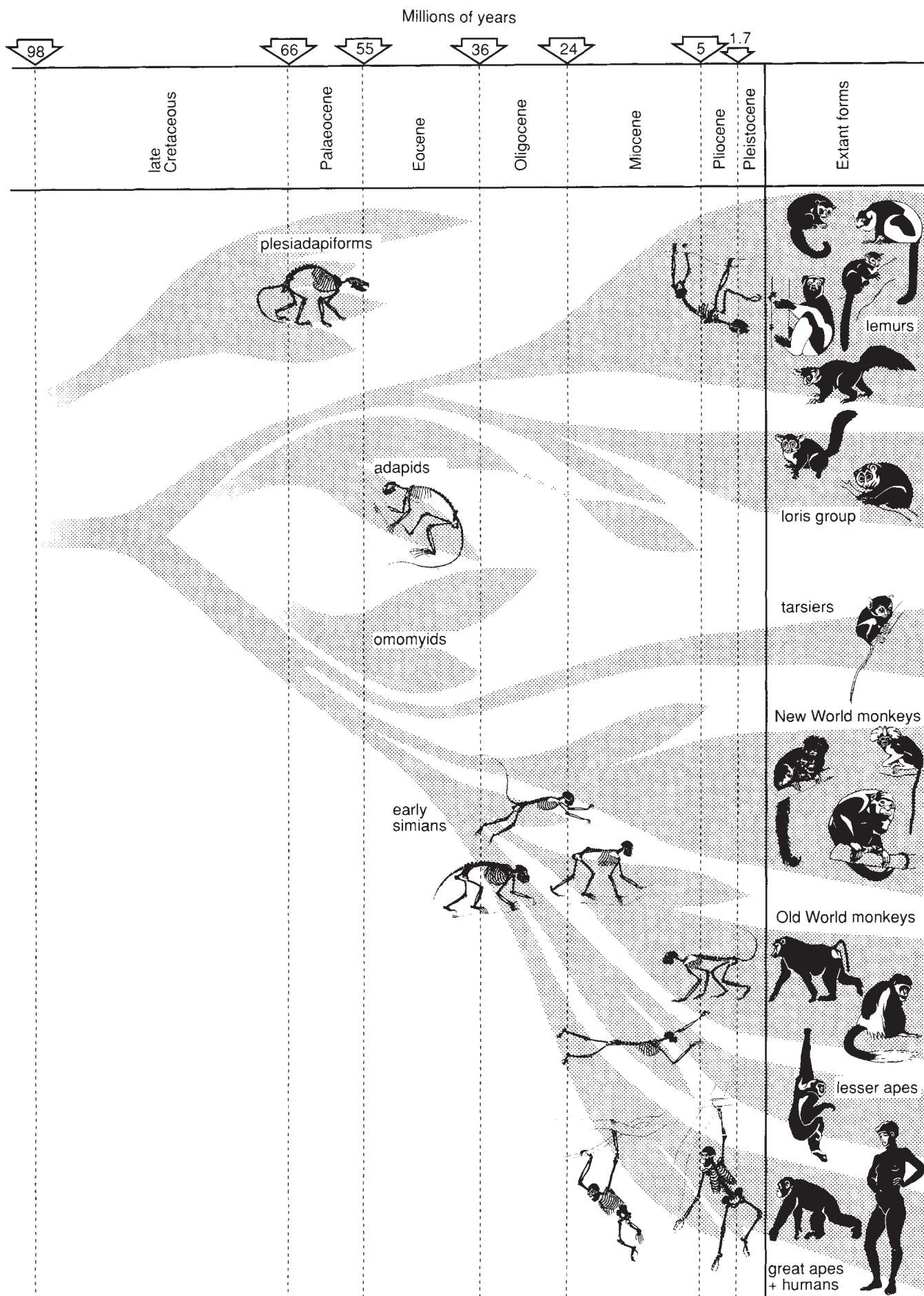
In 1979, when the last comprehensive review of fossil primates¹ was published, it seemed that broad agreement might be emerging with respect to the early evolutionary history of the primates (Fig. 1). It was almost universally accepted that the primates originated towards the close of the Cretaceous, just over 65 million years ago, and an overwhelming majority of primate palaeontologists regarded the Plesiadapiformes as members of the first major adaptive radiation from ancestral primates. Although these 'archaic primates' increasingly came to be seen

as a side branch, with no more than an early connection to the main radiation that led to modern primates, few doubted that the Plesiadapiformes were the exclusive sister group of primates of modern aspect ('euprimates'). There was also widespread, if not universal, agreement that the earliest known fossil primates of modern aspect from the Eocene fall into two distinct groups corresponding to the divergence of two main groups of living primates. The Adapidae were often regarded as relatives of the modern strepsirrhine primates, whereas the Omomyidae were believed by many authors to be specifically allied to modern tarsiers. For the growing number of authors who accepted the existence of a haplorhine group containing tarsiers and simians, the omomyids almost automatically came to be seen as a key group close to the ancestry of all modern haplorhines.

Challenges to this picture of primate evolution have arisen in part because of new fossil material. But revised interpretations based on comparative studies and theoretical considerations have also contributed. In fact, exploration of the possible effects of a low sampling level on interpretations of the primate fossil record (Box 1) indicates that radical revision of prevailing views of primate evolution may be inevitable. Assessment of the possible effects of a low sampling level lead to certain predictions: (1) there is still enormous scope for the discovery of new fossil primate species; (2) with increased sampling of the primate fossil record, inferred times of origin are likely to increase, sometimes dramatically; (3) given the difficulties of interpreting fragmentary fossils, newly discovered specimens are likely to lead to repeated radical revision of inferred phylogenetic relationships; (4) the geographical context of primate evolution will also require reappraisal, as earlier times of origin necessarily imply different patterns of proximity between drifting continents.

Relationships of Plesiadapiformes

The revolutionary proposal that some or all plesiadapiforms may be directly related to the modern colugo *Cynocephalus* rather than to primates of modern aspect²⁻⁴ rests on two main lines of evidence. The first stems from a study of postcranial features by Beard⁵. The most striking inference, stimulated by new skeletal material, is that the hand of early Eocene *Phenacolemur* and *Ignacius* (both members of the family Paromomyidae) resembled that of *Cynocephalus*. For both *Phenacolemur* and *Ignacius*, the intermediate digital bones (phalanges) of each finger were reported to be longer than the basal phalanges, as in *Cynocephalus*, possibly indicating adaptation to support webbing between the fingers as part of a patagium (gliding membrane)^{2,6,7}. Further, a cladistic analysis of the fossil evidence (based mainly on postcranial features) led to the conclusion that shared derived features link various plesiadapiforms more closely to colugos than to primates. Skeletons of *Plesiadapis*, however, do not seem to show any gliding adapta-



BOX 1 How significant are gaps in the primate fossil record?

THE fossil record of primate evolution is obviously incomplete, despite major continuing achievements by field palaeontologists. But how large are the gaps? Do we merely need to bridge over a few spaces here and there, or are there in fact yawning chasms? Our effectiveness in sampling past primate species has major implications for interpretations of primate evolution based on the known fossil record.

A crude estimate of our sampling level of the primate fossil record can be obtained relatively simply^{14,105}. About 200 primate species exist today and it is widely accepted that the primates originated at least 65 Myr ago. To make a very rough calculation of the number of fossil primate species that have ever existed, it is necessary to make some assumption about the pattern of expansion of species numbers over time and to have an estimate of the average survival time for individual species. The simplest assumption for expansion of species numbers over time is a uniform rate of increase. If anything, this is likely to underestimate the numbers of extinct species, as it is commonly believed that the rate of speciation is initially high within a given group and then decelerates as available ecological niches become occupied. For fossil mammal species, several studies indicate that the average survival time is about 1 million years¹⁰⁵. Taking all of these elements together, it is possible to infer that about 6,500 extinct primate species preceded the modern fauna.

This estimate can be used to assess our sample of the primate fossil record. Just over a decade ago, one review recognized 64 species of archaic primates and 186 fossil primates of modern aspect¹. These 250 species represent 3.8% of the estimated total of past primate species. If the calculation is restricted to primates of modern aspect, commonly thought to have emerged about 55 Myr ago, the sampling level (186 out of an estimated total of 5,500 species) declines to only 3.4%. The potential implications of such low sampling levels for interpretation of the primate fossil record can be illustrated with a simple tree (Fig. 2). Although random distribution of discovery of fossil species within the tree is in fact unlikely, this does not affect the two main points that can be recognized. Firstly, times of origin for the group as a whole and for subgroups within the tree are likely to be considerably earlier than indicated by first known fossil representatives. It is common practice to date the time of origin of any group from just before the first known fossil representative, but this will obviously be seriously misleading at very low sampling levels. Indeed, some lineages within the tree may remain completely unrepresented in the known fossil record. This leads on to the second point that a very low sampling density may lead us to assign inappropriate ancestral positions to known fragmentary fossils.

Repeated simulations of the effects of different sampling levels for the tree shown in Fig. 2 yield average figures for the earliest fossil species that will be discovered within such a tree by chance and the correction factors required to infer the true time of origin of the group (Fig. 3). Preliminary calculations conducted by S. Tavaré (University of Southern California) have shown that the values derived by such simple simulation (Figs 2 and 3) are quite close to those that would be predicted using

branching process models, although some minor correction may be necessary. Such a simulation approach is admittedly approximate and any real tree is likely to show significant departures from the simple model. In many cases, however, sampling of the fossil record has been even poorer than assumed in the model. For example, primates of modern aspect are particularly well known from the Eocene: 83 of the 186 species mentioned above are derived from that epoch. During this exceptionally warm period, primates, characteristically inhabitants of tropical or subtropical forests, temporarily expanded from the south into the higher latitudes of the northern continents, where the probabilities of fossilization and subsequent fossil discovery were markedly higher. Because of significant geographic expansion, there may well have been a major bulge in species numbers within the primate tree during the Eocene, so a uniform model would underestimate the total number of extinct primate species.

For illustration, the derived correction factors can be applied to the known primate fossil record. With a sampling level of 3.8%, it is necessary to add a correction factor of about 30% to the time of origin indicated by the earliest known species. Hence, if archaic primates are accepted as genuine primates, an inferred time of origin of 65 Myr ago based on the first known representative would be revised to 85 Myr, well back in the Cretaceous. If the calculation is restricted to primates of modern aspect, the sampling level falls to 3.4%, requiring a correction factor of 33%, and a time of origin of 55 Myr ago based on the first known representative would be revised to 72 Myr, also a Cretaceous date.

There is, however, a further reason why the calculations illustrated in Figs 2 and 3 can be expected to underestimate the effects of a low sampling level for fossil primate species. An initial correction of the time of origin of primates of modern aspect based on the data in Fig. 3 yields a figure of 72 Myr. If, however, primates of modern aspect actually originated 72 Myr ago, rather than 55 Myr, the model in Fig. 2 yields a revised total of 7,200 extinct primate species. The sampling level for 186 known species correspondingly drops to 2.5%, which requires a correction factor of 40% instead of 33%, indicating an even earlier date of 77 Myr for the origin of primates. Serial recalculations of this kind eventually lead to a date of about 80 Myr for the origin of primates of modern aspect. A similar series of calculations can be applied to the simian primates, the earliest known fossil representative of which, until very recently, was dated at about 36 Myr, on the Eocene–Oligocene boundary. Application of a similar correction factor to this date would indicate an origin of simians at about 52 Myr, close to the Palaeocene–Eocene boundary.

It should also be noted that such recalculation of times of origin within the primate tree would have secondary implications for calibration of phylogenetic trees based on biomolecular data. Calibration has often been based on the suspect procedure of taking the date of the first known representative of a given group as close to the time of origin. Quite aside from other problems associated with biomolecular trees, it might be advisable to add 40% to all dates of divergence based on calibration derived from properly identified fossil evidence. □

tion. The second main source of evidence is a new, almost complete skull of *Ignacius*. Unlike primates of modern aspect, in which the auditory bulla grows out from the petrosal, the bulla of *Ignacius* can be confidently interpreted as the product of a separate entotympanic element (Fig. 4). As the presence of a petrosal bulla is generally regarded as a defining feature of primates, often stated to apply to plesiadapiforms as well, the entotympanic bulla of *Ignacius* sets paromomyids (at least) apart from primates. It has further been claimed that *Ignacius* specifically resembles colugos because the medial part of the

entotympanic bulla contacts the basioccipital on the underside of the skull, that the bulla in *Plesiadapis* may well be formed from an entotympanic as well, and that *Ignacius*, *Plesiadapis* and colugos also share an unusual secondary regression of the internal carotid system³. The inferred relationship between *Ignacius*, *Plesiadapis* and colugos was later reinforced by a cladistic analysis of 33 cranial features⁸.

Reassessment of the relationships of the Plesiadapiformes has been increasingly hampered by mounting evidence that the group may not be monophyletic. A recent major review, focused on dentitions, divided the Plesiadapiformes into two superfamilies: Microsyopoidea (families Microsyopidae and Palaeacanthonidae) and Plesiadapioidea (families Plesiadapidae, Paromomyidae, Carpolestidae, Saxonellidae and Picrodontidae)⁹. The family Microsyopidae has had a chequered history, being sometimes excluded from the plesiadapiform group^{1,10} and sometimes included^{11,12}. Beard¹³ omitted Microsyopidae, Palaeacanthonidae and Picrodontidae from his analysis because of the lack of fossil evidence. For the remaining families, he proposed a new three-level arrangement for the order Dermoptera with colugos linked first to Paromomyidae, then to a group containing Plesiadapidae, Saxonellidae and Carpolestidae, and

◀ FIG. 1 Outline tree showing the main groups of living primates and one possible interpretation of their relationships. The tree includes fossil forms known from relatively complete skeletons and emphasizes remaining gaps (for example, no fossil lemurs other than recent subfossils; no fossil tarsiers apart from a single lower molar from the Miocene¹¹⁷). Four key groups are discussed with respect to the early evolution of primates: plesiadapiforms; adapids; omomyids; early simians. Because of relatively low fossil sampling levels (Box 1), the time of origin of primates of modern aspect has been set at about 80 Myr ago, rather than at the value of 65 Myr often cited. The time of origin of simian primates has similarly been increased to about 55 Myr. (Adapted from ref. 14; illustration by Lukrezia Bieler-Beerli).

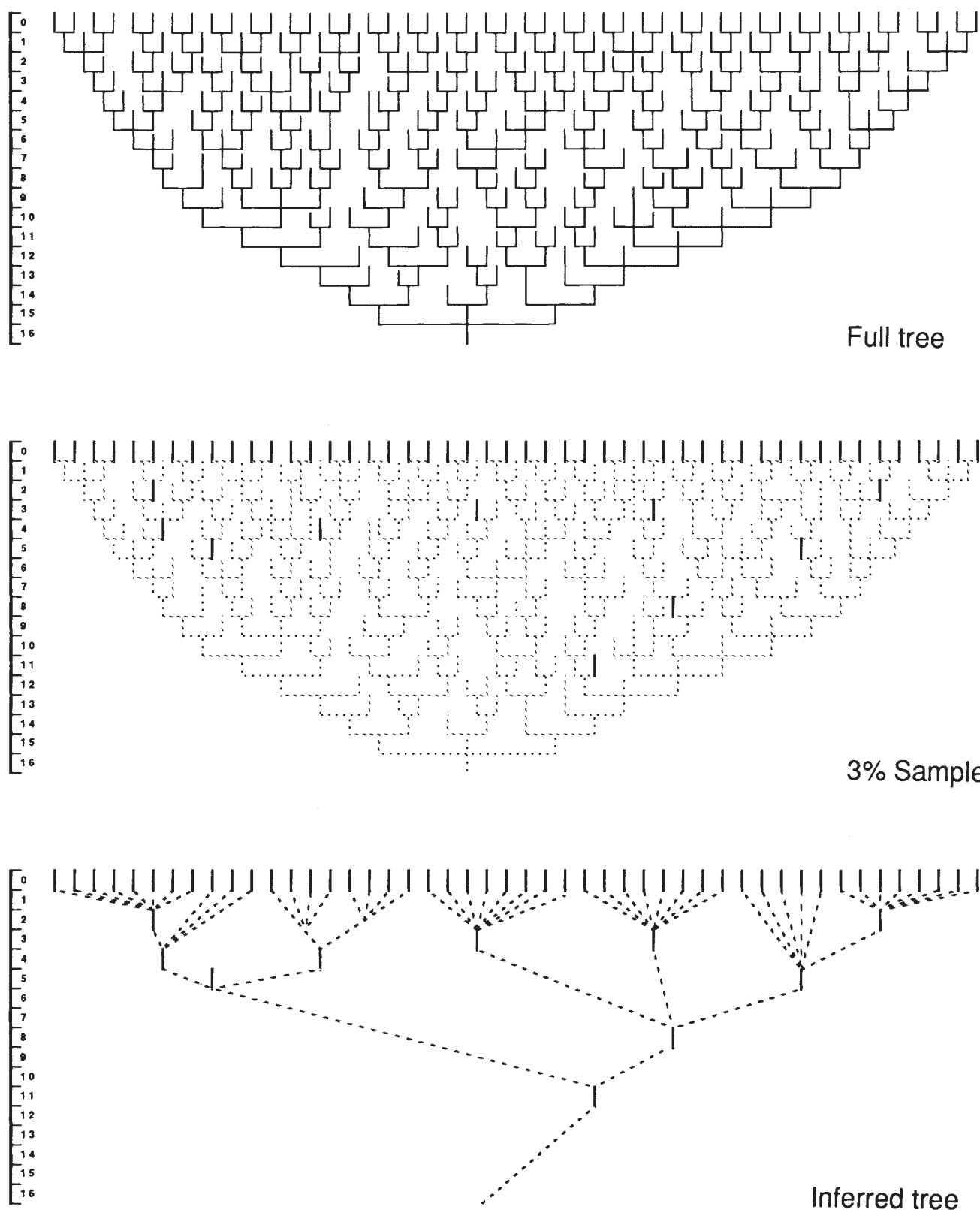


FIG. 2 Full tree: Model branching tree, with progressive expansion of number of species from 1 to 48 over a period of 16 million years. Each species has been given a standard survival time of 1 million years. 3% Sample: A typical example in which a 3% sample of fossil species ($n=10$ species) has been randomly distributed throughout the tree. In this case, the earliest known species is dated at 11 Myr and underestimates the true time of origin of

the group by 5 million years (corresponding to a required correction factor of 45%). Inferred tree: An extreme illustration of the kind of tree that might be reconstructed if major gaps in the fossil record are not acknowledged. The time of origin of the group (based on the first known species) is likely to be seriously underestimated and fossil species may be forced into unrealistic ancestral positions.

finally to Micromomyidae (Fig. 5). The plesiadapiform families included are hence not monophyletic. Beard combined his thus expanded order Dermoptera with primates of modern aspect in the new mirorder 'Primatomorpha'. By contrast, Kay *et al.*⁸ conclude from their cladistic analysis of craniodental features that plesiadapiforms (excluding Microsyopidae) are a monophyletic sister group of colugos and propose that all plesiadapiforms should now be transferred to the Dermoptera^{3,8}. Their analysis indicates no connection between plesiadapiforms and primates, confirming the conclusion from a previous re-examination of evidence¹⁴.

The evidence for a specific link between plesiadapiforms and colugos is open to question. In particular, Krause¹⁵ challenged the inference of gliding adaptations in *Phenacolemur* and *Ignacius* on a number of grounds. His primary criticism is that there are doubts about identification of isolated postcranial bones attributed to *Phenacolemur* and *Ignacius*, particularly the assignment of phalanges to individual digits of hand or foot. As the interpretation of isolated postcranial bones is always difficult, the inference of gliding adaptations in *Phenacolemur* and *Ignacius* is provisional until associated skeletal remains are found. More complete skeletal evidence might, for instance, show whether the atlas vertebra was butterfly shaped for the attachment of the patagium, or whether the ulna was attenuated and united at its lower end with the radius, as in modern *Cynocephalus*. Evidence from the bulla of *Ignacius*, although eliminating one of the main arguments for linking plesiadapiforms to primates, does not necessarily indicate a link with colugos (Fig. 4). Numerous mammals incorporate one or more entotympanics in the bulla and contact between the entotympanic and the basioccipital, presumably reflecting medial expansion of the bulla, could easily arise by convergence. The proposal that a lateral collar-shaped ectotympanic is a shared derived feature of plesiadapiforms and colugos³ is unconvincing as the bulla of colugos is in fact formed almost entirely from the ectotympanic. In contrast to the condition in *Ignacius*, a small rostral entotympanic and an even smaller caudal entotympanic make only minor contributions to the bulla in *Cynocephalus*¹⁶. This feature is linked to a peculiarity of *Cynocephalus*: the eardrum is almost horizontal. This orientation of the eardrum was primitive for mammals and there has been a general trend towards vertical realignment¹⁴. Among modern mammals, the quasi-horizontal condition has been retained only by monotremes and certain lipotyphlan insectivores. As a retained primitive feature of *Cynocephalus*, this would suggest a very early separation from all other eutherian mammals. *Cynocephalus* does show secondary reduction in the internal carotid system within the bulla, associated with increased emphasis on the vertebral arteries for supply of blood to the brain¹⁶ and a similar condition can be inferred for some Paromomyidae and Plesiadapidae, although not for Micromomyidae^{9,13}. But such ontogenetic suppression of a pre-existing system could easily have arisen convergently.

Several other points are pertinent. First, the Palaeocene/Eocene Plagiomenidae have long been regarded as fossil relatives of colugos because of dental resemblances¹⁷. But new basicranial evidence revealed that *Plagiomena* greatly differs from *Cynocephalus*, indicating that the dental similarity is convergent¹⁸. This provides yet another example of the danger of basing inferred relationships exclusively on dental evidence. More recently, *Dermotherium* from the late Eocene of Thailand has been identified as a direct relative of modern colugos¹⁹. Although the evidence is restricted to a lower jaw fragment containing only two molars, the resemblance to *Cynocephalus* is impressive and may well indicate that the colugos existed as a distinct lineage at least by the late Eocene. *Dermotherium* seems to rule out the Plagiomenidae¹⁷ or the early Eocene Placentidentidae of Europe²⁰ as relatives of *Cynocephalus*.

Curiously, little has been said about possible dental evidence linking plesiadapiforms to colugos. For example, the cladistic

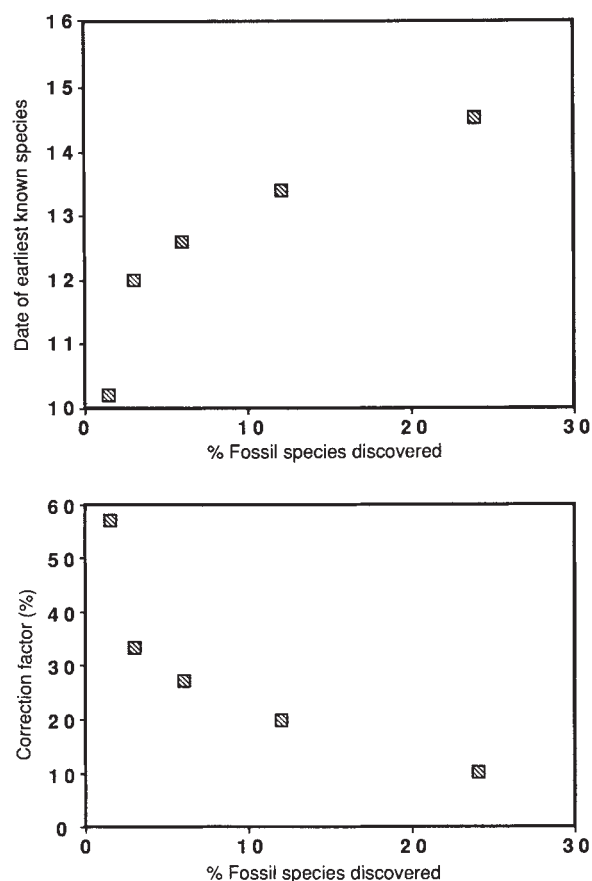
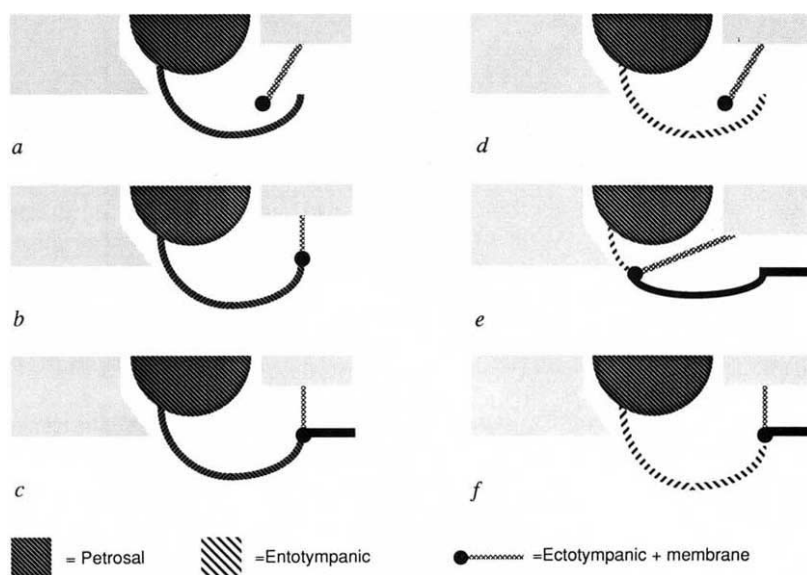


FIG. 3 Upper curve: Plots of average values obtained from 25 simple simulations of various sampling densities in the fossil record (1.5%, 3%, 6%, 12%, 24%), taking the tree in Fig. 2. Relative to a true origin at 16 Myr ago, the date of the earliest known species is close to 10 Myr at a sampling level of 1.5% and is below 15 Myr even with a sampling level of 24%. Lower curve: Correction factors (% value to be added) for such simple trees can be calculated for inferring the actual time of origin from the date of the first known fossil species at different sampling densities of the fossil record.

analysis of Kay *et al.* included no dental evidence⁸. Supposed resemblances in the molar teeth between plesiadapiforms and primates of modern aspect provided the main basis for inferring a phylogenetic link, although this evidence has been questioned¹⁴. Surely, if plesiadapiforms are linked to colugos instead of primates, we should expect plesiadapiforms to show some dental resemblance to *Cynocephalus*. Modern colugos in fact have a highly unusual dentition, with comb-like crowns on the procumbent lower incisors and serrated upper incisors and upper and lower canines. The upper molar teeth are essentially triangular and bear three main cusps, presumably a primitive feature, and all the cheek teeth possess tall, sharp cusps adapted for shearing. *Cynocephalus* reportedly feeds primarily on leaves, but the dietary function of the comb-like lower incisors has yet to be explained. Comparison between *Cynocephalus* and Paromomyidae merely shows that they are dentally very different¹⁹. It is, of course, possible that the dental features of *Cynocephalus* have become so specialized that its relationships are no longer evident⁶. For instance, Beard proposes a *Nannopithecus*-fold on the upper molars as a shared derived feature characterizing the 'Primatomorpha' and then infers that it disappeared without trace from colugos through evolutionary reversal¹³. This dental feature is, in fact, lacking from the earliest known fossil primates of modern aspect, so it is more likely to have arisen by convergence even within primates^{8,14}. In sum, plesiadapiforms, formerly 'dental primates', are no more than 'postcranial colugos'.

FIG. 4 Schematic transverse sections of the auditory bulla in primates (left) and non-primates (right). In each case, the skull midline is on the left and the outer skull margin is on the right. In all living primates, the bulla wall is formed from the petrosal, but the relationship between the ectotympanic ring and the bulla varies between groups: enclosed in lemurs (*a*); fused to the outer margin in the lorises group and in New World monkeys (*b*); fused to the outer margin and extended to form a tube in tarsiers and Old World simians (*c*). In tree shrews (*d*), the bulla wall is formed from an entotympanic and the ectotympanic ring is enclosed. In colugos (*e*), the bulla wall is formed mainly from the ectotympanic, with a small entotympanic contribution. Note the almost horizontal orientation of the tympanic membrane (eardrum). In the plesiadapiform *Ignacius* (*f*), the bulla wall is formed mainly from an entotympanic, with the ectotympanic fused to the outer margin and extended to form a tube. There is convergence in general form (but not in constitution) between tree-shrews and lemurs (*d* versus *a*) and between plesiadapiforms and tarsiers/Old World simians (*f* versus *c*).



If colugos and plesiadapiforms together constitute the sister group of primates, all should possess shared derived features. Beard's cladistic analysis¹³, based mainly on postcranial traits, seems to demonstrate the existence of a special set of adaptations for vertical postures and climbing⁶. By contrast, the analysis of cranial features by Kay *et al.*⁸ leads to the incompatible conclusion that primates are linked not to plesiadapiforms and colugos but to tree-shrews and bats. Moreover, in addition to the horizontal orientation of its eardrum (Fig. 4), *Cynocephalus* shows several features ignored in both analyses that would be highly unusual in any direct relative of primates. The brain, for instance, is remarkably small, with dominant olfactory components and diminutive cerebral hemispheres (Fig. 6). Unfortunately, a report on relative brain size in *Cynocephalus*²¹ cited an immature body weight of only 810 g. Taking a revised adult body weight of 1.4 kg and a cranial capacity of 6.9 ml, the index of cranial capacity (ICC)¹⁴ is only 1.5, on the boundary between 'basal' and 'advanced' insectivores. This value is well below the lower limit for modern primates (2.4) and is instead close to the value of 1.4 for the Eocene primate *Adapis*, which lived some 40 million years ago (Fig. 6). As a further strikingly primitive feature, although *Cynocephalus* has only a single neonate, it is apparently extremely poorly developed (altricial) and has been described as 'marsupial-like'²². Modern primates are uniformly characterized by neonates with well defined precocial features¹⁴.

Inference of a link between primates, plesiadapiforms and colugos is connected with a recent revival of the concept of the superorder Archonta, originally grouping primates, colugos, bats, tree-shrews and elephant-shrews²³. Although the terrestrial elephant-shrews are now generally excluded, there have been repeated proposals of relationships among some or all of the remaining forms, which typically have some connection with arboreal habits. Proposed inclusion of tree-shrews in the order Primates has a long history, but there is no convincing evidence of shared derived features and it is preferable to allocate them to their own order Scandentia^{14,24}. MacKenna's proposal²⁵ that bats, colugos, tree-shrews and primates together form a phylogenetic unit was later supported on the basis of postcranial evidence²⁶, although the status of the bats remained uncertain, whereas immunological evidence apparently linked colugos, tree-shrews and primates to the exclusion of bats²⁷. On morphological grounds, Shoshani identified a phylogenetic association between bats, colugos, tree-shrews and primates²⁸ and Novacek cited limited support for a concept of Archonta in which tree-shrews are allied with primates and colugos with bats²⁹⁻³¹. By

contrast, an analysis of amino-acid sequences in a tandem alignment of up to seven proteins^{32,33} linked bats and tree-shrews with insectivores and carnivores and associated primates with lagomorphs and rodents.

In a new variation on the archontan theme, Pettigrew³⁴ reported evidence from the visual system (pattern of retinotectal projection) indicating that the fruit-bats (Megachiroptera), but not other bats (Microchiroptera), are directly related to primates. Following a cladistic analysis of 24 neural characters, this proposal was then expanded into the 'flying primate hypothesis' linking primates, fruit-bats, colugos and tree-shrews (in that order)³⁵, with paromomyids being included as purported relatives of colugos in due course³⁶. This hypothesis has been opposed because it conflicts with extensive morphological evidence indicating a monophyletic origin for bats^{31,37,38}. It has also emerged that the pattern of retinotectal projection in the fruit-bat *Rousettus aegyptiacus* is not primate-like but follows the general mammalian scheme³⁹, suggesting that a primate-like pattern may have evolved convergently in some fruit-bats (such as *Pteropus*). Further, a sequence analysis of ϵ -globin genes for 17 mammal species⁴⁰ produced very convincing evidence for a monophyletic origin of bats and linked tree-shrews to lagomorphs rather than to primates. Monophyly of bats was independently confirmed by an analysis of the mitochondrial cytochrome oxidase subunit II gene⁴¹. This analysis also groups *Tupaia* and *Cynocephalus* (but not bats) with primates, apparently providing support for a curtailed archontan assemblage. In fact, *Cynocephalus* consistently emerged as closer to primates than *Tupaia*. As the comparisons include only nine placental mammal species, however, these results must be regarded as tentative.

Overall, the superordinal relationships of primates remain unclear. Even authors who accept the existence of a group 'Archonta' generally do not agree on its composition or internal arrangement and no consistent picture is provided by molecular comparisons⁴². All such interpretations are based on undoubted morphological similarities between bats, colugos, tree-shrews and primates, but their interpretation is questionable. Most authors assume that the ancestral placental mammals were terrestrial and accordingly interpret any arboreal adaptations of 'archontans' as shared derived features indicating a specific common ancestry. If, however, the ancestral placental mammals were at least partially arboreal in habits, the similarities shared by colugos, bats, tree-shrews and primates can be interpreted as a mixture of primitive retentions and convergent adaptations¹⁴. In this respect, an important new development has been

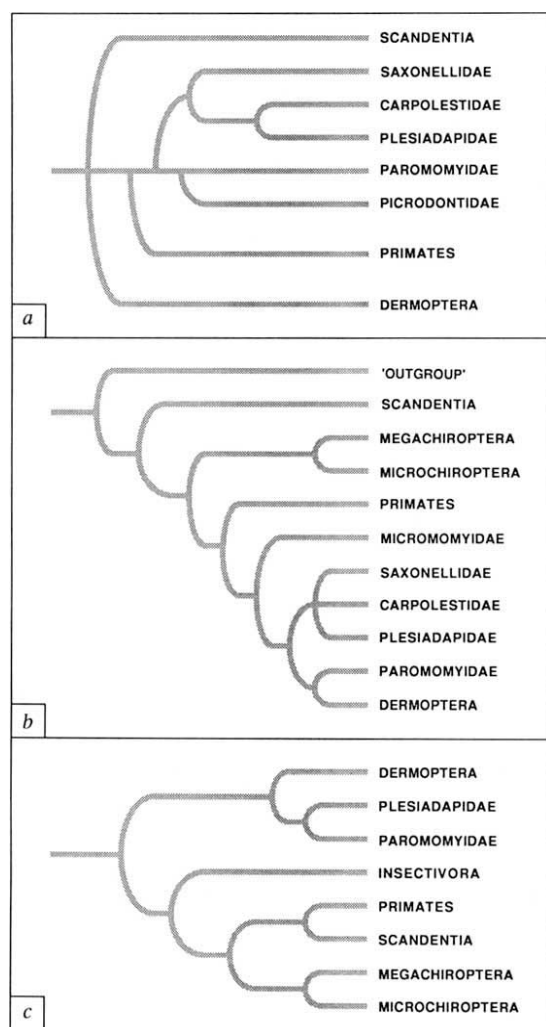


FIG. 5 Alternative interpretations of the relationships between dermopterans, plesiadapiforms and primates, following: a, Szalay and Delson¹; b, Beard¹³; and c, Kay *et al.*⁸.

the description of the postcranial skeleton of the Jurassic eupantothere *Henkelotherium*⁴³, which is surprisingly advanced and shows several adaptations for arboreal life. Therian mammals (marsupials and placentals) are probably derived from a eupantothere stock, so this new evidence strengthens the inference that marsupials and placentals both had an arboreal origin.

Affinities of the Adapidae

New discoveries of fossil adapids ('lemuroids') have included important finds both in North America and in Europe. In North America, where all or most forms can be allocated to the subfamily Notharctinae, the main advance has been a significant increase in our knowledge of the early genera *Cantius* and *Pelycodus*. The later forms, *Notharctus* and *Smilodectes*, have long been documented by some of the best material available for fossil primates^{1,44-46}, but the earlier North American adapids were known essentially from isolated teeth and jaw fragments. Indeed, well preserved skulls of these earlier representatives have yet to be discovered. A study of more than 100 new foot elements from 7 species of *Cantius* and (?) *Pelycodus*⁴⁷ has now partly improved our knowledge. Not surprisingly, *Notharctus* and *Smilodectes* are apparently derived in comparison with *Cantius*, although only in subtle details. Overall, these new finds fit expectation quite closely. It is interesting, however, that the tarsal bones changed relatively little during the early radiation of *Cantius* and (?) *Pelycodus*, although the dentition changed

quite markedly over the 5-million-year period concerned. This is a particularly clear example of mosaic evolution. Comparison of foot bones of *Cantius*, *Notharctus* and *Smilodectes* with those of modern lemurs and lorises provides little evidence of a specific link between the Eocene forms and modern strepsirrhine primates. The only features that may indicate a specific link between all adapids and modern strepsirrhines are morphological features of the ankle region (sloping talo-fibular facet; lateral groove for the flexor hallucis longus; ventral extension of the navicular facet for the cuboid). Such features have been cited⁴⁸ in support of the interpretation that the adapids are the sister group of modern lemurs and lorises, and this inferred link has been reinforced with the suggestion that adapids and strepsirrhines show shared derived features of the hand and wrist associated with an emphasis on hand postures involving ulnar deviation⁴⁹. A link between adapids and modern strepsirrhines has also been proposed on the basis of resemblances in the incisor-canine complex⁵⁰. Taken together, these various lines of evidence provide suggestive but by no means conclusive indicators of a connection between adapids and strepsirrhines.

The main surprises with respect to the adaptive radiation of the adapids have come from recent studies of European forms. The best-known genera with respect to craniodental remains continue to be *Adapis* and *Leptadapis*, still mainly represented by numerous well preserved skulls discovered long ago in late Eocene deposits of France. Although several species were originally identified⁵¹, most recent authors have recognized just the two species *Adapis parisiensis* and *Leptadapis magnus*. A comprehensive re-examination of these skulls has now convincingly shown that quantitative variation among the skulls attributed to either species greatly exceeds that found within any modern primate species, even exceeding the level of variation across closely related genera for some features^{52,53}. Excessive variation in dental dimensions of specimens allocated to '*Adapis parisiensis*' had already been noted by Gingerich^{54,55} and it now seems highly likely that this taxon and '*Leptadapis magnus*' in fact group a number of species together⁵⁶. An earlier suggestion^{57,58} that quantitative differences among skulls of *Adapis* or *Leptadapis* reflect sexual dimorphism must now be re-examined. The new findings do not necessarily show that sexual dimorphism was lacking in *Adapis* and *Leptadapis*, but such sexual differences must now be sought within the individual species identifiable on the basis of Lanèque's analyses. For the

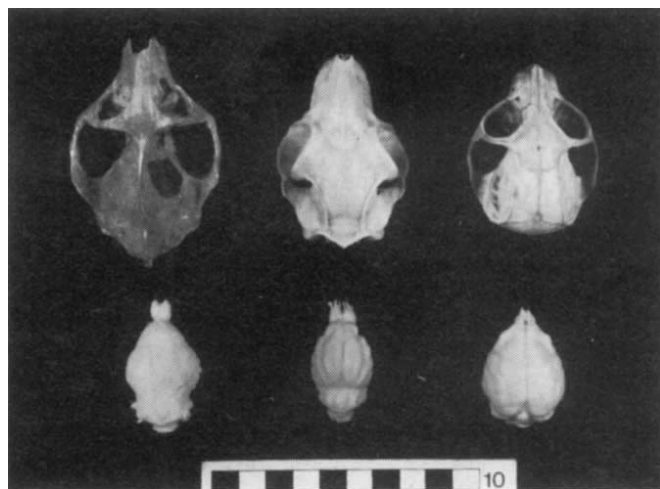


FIG. 6 Skulls and endocasts (casts of the brain cavity) of *Adapis parisiensis* (left), *Cynocephalus variegatus* (centre) and *Perodicticus potto* (right). In both the Eocene *Adapis* and the modern *Cynocephalus* the endocast is very small relative to body size and the neocortex is diminutive, as is indicated by the dorsal exposure of the olfactory bulbs and the cerebellum. Note also the laterally facing orbits, the lack of a postorbital bar and the wide separation of the orbits in *Cynocephalus*, contrasting starkly with the typical primate pattern in *Adapis* and *Perodicticus*.

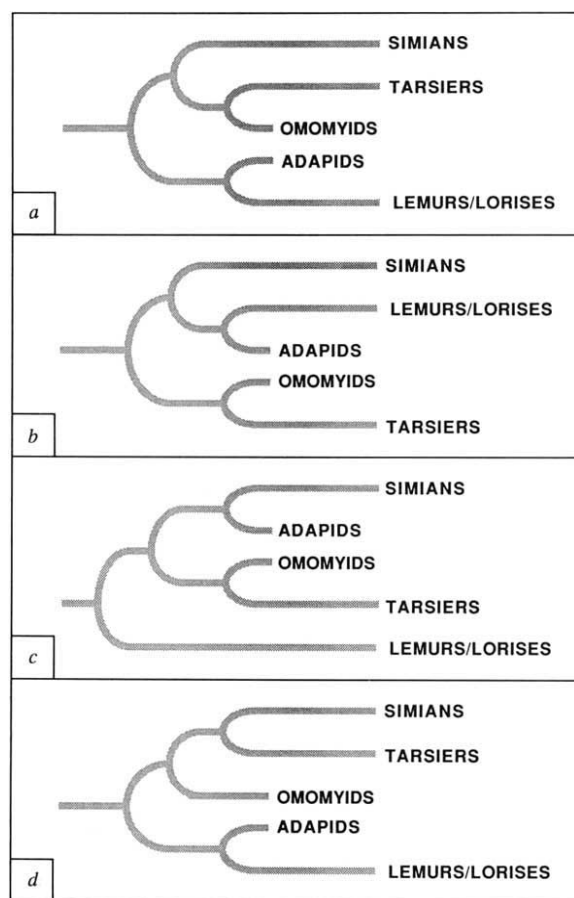


FIG. 7 Four alternative interpretations of the relationships between adapids, omomyids and living primates. *a*, The omomyophile hypothesis⁷⁸, with simians deriving from an omomyid-like stock; *b*, one version of the adapid hypothesis (the lemurphile hypothesis⁷⁸), with both simians and strepsirrhine primates originating from an adapid-like stock; *c*, a second version of the adapid hypothesis (the lemurphobe hypothesis), with simians originating from an adapid-like stock, whereas strepsirrhine primates had a distinct origin; *d*, the tarsiphile hypothesis⁷⁸, with simians deriving from a tarsier-like ancestor.

time being, the only remaining evidence for sexual dimorphism in Eocene primates is that reported for the upper canine teeth of *Notharctus ventriculus*⁵⁹. In that case, the specimens concerned came from the same stratigraphic horizon and showed two different canine sizes despite the fact the molar teeth were all very similar in size and shape, so it is less likely that interspecific differences were confused with differences between the sexes.

In contrast to information on the North American representatives, our understanding of European adapids has long been severely constrained by the lack of associated postcranial remains. Isolated postcranial bones have been attributed to *Adapis* and *Leptadapis* on the grounds of size and primate-like features, but calcanei (heel-bones) attributed to these two genera lack the relative elongation of the anterior segment that is otherwise universally found among fossil and living prosimian primates^{14,60,61}. It is possible, as concluded from the detailed study conducted by Dagosto⁶², that *Adapis* and *Leptadapis* exhibited extensive secondary specialization of the postcranial skeleton, accounting for departures from the typical primate pattern. Additional evidence of the puzzling nature of postcranial elements attributed to *Adapis* and *Leptadapis* is provided by a recent study⁶³ in which body weight of fossil primates was estimated from dimensions of the tarsal bones. Overall, there is relatively good agreement between these postcranial estimates and those derived from skulls and/or teeth. There is one surprise

in that *Adapis* skull measurements consistently indicate a body weight of at least 2 kg (ref. 14), whereas postcranial elements attributed to this genus indicate a body weight of only 1 kg; for *Leptadapis*, both skulls and postcranial elements indicate the same body weight of about 8.5 kg. It is possible that this disparity will disappear when the different species currently subsumed in '*Adapis parisiensis*' have been clearly separated, but only the discovery of associated cranial and postcranial remains would settle the issue clearly.

In recent years, postcranial remains of a number of other European adapids have been reported. Most finds have come from the Messel deposits of southern Germany, which have yielded five specimens⁶⁴⁻⁶⁸. Unfortunately, the specimens are all incomplete and three of them cannot be identified even to genus level, because the skull and teeth have not been preserved, and they have hence been recognized simply as adapids of some kind. One specimen, however, consists of the anterior part of the body including a crushed skull and has been identified as *Europolemur koenigswaldi*, while a second specimen has been identified as the rear end of the same species. These specimens indicate that *Europolemur* was a relatively small-bodied arboreal climber and leaper with extremities adapted for grasping⁶⁶. Another crucial new specimen from a different site is the crushed skull and skeleton of *Pronycticebus neglectus* from the Geiseltal deposits of eastern Germany⁶⁹. The genus *Pronycticebus* was previously documented only by a single skull from France and study of the German skeleton indicates that this primate, which probably had a body weight close to 750 g, was an arboreal climber and leaper with an insectivorous/frugivorous diet. Finally, recent excavations in northern Spain have yielded more than 500 specimens representing most of the postcranial skeleton attributed to the hitherto poorly known genus *Anchomomys*⁷⁰. This genus is of particular significance because its body size (about 120 g) is much smaller than that of other adapids. Interestingly, the postcranial elements of all of these new adapid finds are closely similar to those of the North American notharctines and unlike those attributed to *Adapis* and *Leptadapis*.

One possible interpretation of the new finds is that at least some postcranial elements attributed to *Adapis* and *Leptadapis* do not in fact belong to those genera. By contrast, if the allocations are correct, it follows that *Adapis* and *Leptadapis* are quite distinct from most other known adapids. Gingerich⁷¹ originally distinguished between an '*Adapis* group' and a '*Protoadapis* group' in Europe, although he did not identify a major separation between them. Later, citing dental characteristics and postcranial features, Franzen⁶⁶ proposed that the European adapids should be split into two distinct groups: (1) *Adapis*, *Leptadapis*, *Microadapis* and (?) *Caenopithecus*; (2) *Cantius*, *Protoadapis* and *Europolemur*. The second group, characterized (for example) by lower-crowned molars with weakly defined crests, by a transverse oblong shape of the upper molars and by more limited molarization of the premolars, roughly corresponds to the tribe Protoadapini recognized by Szalay and Delson¹. This tribe also includes *Pronycticebus* and the distinctive North American genus *Mahgarita*. Although the Protoadapini are morphologically closer to the North American forms and are represented in Europe throughout the Eocene, *Adapis*, *Leptadapis*, *Microadapis* and *Caenopithecus* appear abruptly towards the end of the middle Eocene⁷². For now, it is probably best to regard the *Adapis* group as an aberrant sideline that may or may not be a part of a monophyletic radiation of adapids. By contrast, the main group is of particular interest because of suggestions that the origins of simians may be traced specifically to the Protoadapini (Box 2).

It is also significant that a single lower molar tooth (M_2) from a presumed middle Eocene locality in South-East Asia has been identified as an adapid⁷³. Confirmation of this identification would greatly expand the known geographical range of Eocene adapids and could lead to new interpretations of early primate evolution.

BOX 2 What was the origin of the simian primates?

Four main hypotheses have been proposed to explain simian origins: (1) the tarsier hypothesis, deriving simians from a tarsier-like ancestor¹⁰⁶; (2) the omomyid hypothesis, deriving simians from an omomyid stock¹⁰⁷; (3) the adapid hypothesis, deriving simians from an adapid stock¹⁰⁸; (4) the pre-simian hypothesis, deriving simians from hypothetical ancestors in Africa during the early Tertiary¹⁰⁹ or in Gondwanaland during the upper Cretaceous¹¹⁰. A widely favoured interpretation has been that omomyids gave rise to both tarsiers and simians (Fig. 6a). That interpretation has been increasingly challenged, however, and several authors now prefer some form of the adapid hypothesis instead. One version is that adapids gave rise to both simians and strepsirrhine primates (Fig. 6b)^{46,71,79}, but this is in direct conflict with abundant evidence from living forms that tarsiers and simians form a monophyletic group. An alternative version is that adapids exclusively gave rise to simians (Fig. 6c)^{81,96,111–114}. Another possible interpretation⁷⁸, according to which the omomyids branched away before the separation between tarsiers and simians (Fig. 6d), in effect amounts to a revival of the tarsier hypothesis.

With respect to the known fossil record, the discussion boils down to the question whether omomyids or adapids are most likely to have given rise to simians. There is little if any direct evidence for a link between omomyids and simians (Fig. 6a or d); instead there has generally been an indirect chain of argument that tarsiers are related to omomyids, that tarsiers are also related to simians and that omomyids must thus be related to simians. If tarsiers are not, in fact, related to omomyids, the argument for an omomyid origin of simians virtually collapses.

The alternative proposal that simians are derived from adapids has been backed on the basis of a seemingly impressive list of similarities, including increased body size, vertically implanted spatulate incisors, large interlocking canines with a honing relationship between the upper canine and the anterior lower premolar, sexual dimorphism in body size and canine size, a quadratic form of the molars, fusion of the anterior junction (symphysis) between the two halves of the lower jaw, an annular form of the ectotympanic, unfused tibia and fibula and lack of elongation of the calcaneum and navicular^{46,79,111}. Some of these characters can be discounted because they are likely to be primitive for primates generally (for example, large canines; annular ectotympanic; unfused tibia and fibula; non-elongated calcaneum and navicular), whereas others have developed so often during mammalian evolution that they carry little weight (for example, increased body size; quadratic molars). Further, the evidence for sexual dimorphism in adapids now needs careful re-examination and such dimorphism is in any case neither unique to nor ubiquitously present in simian primates.

Spatulate incisors and fusion of the lower jaw symphysis have all arisen repeatedly in the evolution of primates and other mammals. Indeed, within the primates these features were developed independently in various

large-bodied subfossil lemurs and similarities in jaws and teeth between *Archaeolemur* and Old World monkeys provide one of the most striking cases of convergent evolution known for primates. Further, fusion of the lower jaw is not found in all adapids. It was lacking in the earliest adapids (*Cantius*; *Pelycodus*) and was absent even in some later forms, such as *Smilodectes*^{115,116} and *Pronycticebus*⁶⁹, so this feature must have arisen several times independently within the Adapidae^{14,46,66}. Symphyseal fusion occurred late in development in *Notharctus*; only in *Adapis*, *Leptadapis* and (presumably) *Mahgarita* did fusion occur early in life as in simians. Hence, if simians are to be derived from adapids on grounds of symphyseal fusion, they must presumably be derived from one of these genera. First, however, *Adapis*, *Leptadapis* and *Mahgarita* are all late Eocene forms, most probably too recent for simian ancestry. Second, it now seems that in the earliest yet known simians, such as late Eocene *Catopithecus*⁹⁶ and *Arsinoea*⁹⁷, the symphysis may not have been fused, in which case symphyseal fusion must have developed after the origin of these simian primates.

Rasmussen and Simons¹¹³ narrowed the possible source of simians among adapids to the tribe Protoadapini, and Rasmussen¹¹⁴ subsequently noted additional specific resemblances between the North American genus *Mahgarita* and simians. These included: large promontory canal; stapedial canal reduced or absent; pneumatized petromastoid region; lateral transverse intrabullar septum; probably no free ectotympanic ring; mandible firmly fused and robust; mandibular symphysis with a transverse torus; short, deep maxilla; detailed similarities in occlusal features of upper molars and other teeth. Although these similarities are intriguing, they can only be of phylogenetic significance if simians are derived directly from *Mahgarita* or some similar ancestral form. One must bear in mind the possibility that sharing of a number of similarities between certain adapids and simians may be attributable to convergence.

Derivation of simians from either adapids or omomyids remains problematic because the known fossil skulls uniformly lack any indication of postorbital closure by a bony partition and the frontal bones remain unfused into adult life. Similarly, known adapid skulls show no sign of the early marked increase in relative brain size that distinguishes simians from prosimians overall and is already established in *Aegyptopithecus*, although such an increase is in fact identifiable in Eocene omomyids^{14,46}. Finally, no known fossil skulls before the Fayum simians show the loss of the median gap between the upper incisors that characterizes tarsiers and simians and would indicate loss of the rhinarium (haplorhine condition). In short, even if simians can be derived from adapids or omomyids, vital transitional evidence is still lacking. In other words, the pre-simian hypothesis, tracing simians to an as yet undocumented ancestral stock in the southern continents, is still a viable option. □

Affinities of the Omomyidae

The status of the early Tertiary Omomyidae ('tarsioids') is arguably the most controversial issue concerning the evolutionary tree of primates and seems set to remain so for the foreseeable future. Not long ago, it was widely accepted that omomyids are directly linked to the ancestry of tarsiers. Similarities between tarsiers and at least some omomyids exist in the bell-shaped upper dental arcade, unusual foreshortening of the lower jaw associated with a reduction in number of teeth relative to the upper jaw, a tubular meatus in the bulla and adaptation of the postcranial skeleton for leaping. In addition, omomyids do seem to resemble both tarsiers and simians in possessing a short facial skull associated with a relatively small nasal fossa⁷⁴, and there is some evidence of limited reduction in the size of the olfactory bulbs¹⁴. Although it was suggested that omomyids resemble tarsiers and simians in possessing an enlarged promontory artery that enters the bulla posteromedially⁷⁵, *Necrolemur* apparently possessed a relatively large stapedial artery and the medial entry into the bulla may in fact be primitive for mammals^{76–78}. Rosenberger⁷⁴ has noted further similarities in the basicranium between some omomyids and tarsiers, including a flexed basi-cranial axis, encroachment of the bulla on the pterygoid fossa associated with extensive laminar confluence between pterygoid and bulla, and a deeply guttered, slot-like shape of the glenoid fossa. But there has always been an undercurrent of doubt about

the inference of a link between omomyids and tarsiers and interpretation of the fossil record is hampered by the fact that modern tarsiers have molar teeth that seem extremely primitive. The upper molars have a simple three-cusped pattern, and the paraconid cusp (commonly lost in fossil and living primates of modern aspect) is still present in the lower molars. Further, while it is true that omomyids resemble tarsiers in possessing paraconids on their lower molars, this is no more than retention of a primitive feature. Hence, it is difficult to link omomyids to either tarsiers or simians through dental morphology.

One alternative view is that the omomyids are, indeed, related to tarsiers and that both are appropriately allocated to the infraorder Tarsiiformes, but that the latter has no connection with simians. An extreme version of this interpretation was advocated by Gingerich¹¹, who allied Tarsiiformes with Plesiadapiformes in the suborder Plesitarsiiformes because of apparent similarities in the structure of the bulla and in the anterior lower dentition. All other primates of modern aspect (lemurs, lorises and simians) were combined in a second suborder Simiolemuriformes. This arrangement, however, requires convergent evolution of many features present in all living and fossil primates of modern aspect (including tarsiiforms) but lacking in plesiadapiforms. Gingerich subsequently accepted a model of primate evolution incorporating three adaptive grades: Plesiadapiformes, Prosimii (including lemurs, lorises, tarsiers,

Adapidae and Omomyidae) and Anthroidea¹² and later retracted his suggestion of a link between plesiadapiforms and tarsiforms⁷⁹. His revised model, however, incorporated no direct connection between tarsiforms and simians. This left the way open for the less extreme proposal that the origin of simians within the prosimian grade is to be traced not to omomyids but to adapids (Box 2). In fact, both of these alternative proposals are in direct conflict with many similarities, several of which are unique, linking modern tarsiers to simians^{14,80}. It is, of course, conceivable that some similarities have arisen convergently. For instance, developmental evidence seems to indicate that the almost complete bony eye-socket of tarsiers and simians (a feature found in no other mammals) may have arisen independently⁸¹. Nevertheless the sheer numbers of apparent derived features in independent systems shared by tarsiers and simians render convergence in all of them extremely unlikely.

A different challenge to the consensus view linking omomyids, tarsiers and simians comes from the view that there is no specific link between omomyids and tarsiers^{77,82}. One feature linking modern tarsiers and simians (haplorhine primates) is the complete loss of the rhinarium (an area of naked skin surrounding the nostrils). This derived feature, which is highly unusual among mammals, is correlated with closure of a median gap between the upper incisors that occurs in association with the rhinarium in lemurs and lorises (strepsirhine primates). It is now known that the European omomyids *Necrolemur* and *Pseudoloris* still had a gap between the upper incisors^{77,80}. It is therefore possible that a rhinarium was still present in omomyids, although its loss could have preceded closure of the gap between the incisors. In any event, broad reviews of the evidence^{77,78,80,83} have indicated fairly clearly that the relationship between tarsiers and simians is probably closer than that between tarsiers and omomyids. Indeed, Schmid⁸² has suggested that omomyids should be classified together with lemurs, lorises and adapids in the suborder Strepsirhini.

By contrast, there have been a number of important developments concerning omomyids in recent years, and the authors involved have tended to support a link between omomyids and tarsiers. One interesting development has been identification of a reasonably complete skull of *Microchoerus* from the Eocene of France⁸⁴. This skull had long been allocated to the closely related genus *Necrolemur*, which is documented by several skulls. The orbits of *Microchoerus* are somewhat larger than in *Necrolemur* and the upper dental arcade does not clearly show the bell shape typically found in other European omomyids. Postcranially, the European omomyids *Necrolemur* and *Microchoerus* differ from the North American forms in having far greater elongation of the anterior segment of the calcaneus⁸⁵, representing a development toward the condition found in modern bushbabies and tarsiers. It has also been claimed that the European forms *Necrolemur* and *Nannopithecus* specifically resemble modern tarsiers in showing distal fusion between the tibia and fibula, although the evidence for *Nannopithecus* has not stood up to closer examination and that for *Necrolemur* is questionable (P. Schmid, personal communication). Rosenberger and Dagosto⁸⁴, however, emphasize apparent derived features of the jaw joint and skull base shared by tarsiers and the European omomyids (Microchoerinae), concluding that there is a specific link between them and that the simians arose from less specialized North American omomyids (anatomorphines or omomyines). This follows the general trend of inferring a specific link between tarsiers and microchoerines, which were in fact included in the Family Tarsiidae by Simons⁸⁶.

More dramatic was the discovery of four new skulls of the North American omomyid *Shoshonius*, previously documented only by fragmentary jaws and teeth^{87,88}. Previously, the skull of Eocene omomyids from North America was known only from an incomplete cranial specimen of *Tetonius*, so the new skulls are in any case of great importance for our understanding of early omomyids. According to Beard *et al.*⁸⁷, however, the skulls

of *Shoshonius* are of special significance because some features, not found in other omomyids, indicate a direct relationship to modern *Tarsius*. This would simultaneously confirm the derivation of tarsiers from omomyid stock, although from omomyines rather than microchoerines, and would push the time of origin of omomyids well back into the Palaeocene. *Shoshonius* resembles tarsiers in having relatively very large orbits, in having a very short snout and in a number of specific features of the ear region. These resemblances are all rather limited, however, and could have arisen through convergent evolution. The possibility of a specific relationship between *Shoshonius* and *Tarsius* could be tested with postcranial evidence. If, for instance, *Shoshonius* were found to show fusion of the tibia and fibula and the same peculiar adaptation of the ankle region as in *Tarsius*⁸⁹, this would provide more convincing evidence of a specific link between these two genera. Postcranial elements and additional skull material of *Shoshonius* have been found, so more information should be forthcoming. At present, the interpretation of omomyid relationships remains confused, with several possibilities still open. On the basis of isolated cheek teeth, the late Paleocene *Altiasius* from Morocco has been interpreted as an early omomyid⁹⁰, perhaps indicating an African connection.

Early simian primates

Until recently, it was often inferred that simian primates (Anthroidea) first emerged close to the boundary between the Eocene and the Oligocene, just over 35 million years (Myr) ago¹. The earliest known undoubted simians had been recovered from the Fayum deposits in Egypt, the best-documented genus being *Aegyptopithecus*, which exhibits numerous simian features in the skull and postcranial skeletons. A middle Oligocene age was originally accepted for the Fayum deposits, but subsequently the date was revised to lower Oligocene age⁹¹. Although it has now been suggested that the entire sequence of the Fayum deposits should be dated as late Eocene⁹², palaeomagnetic reversal stratigraphy indicates that they straddle the Eocene-Oligocene boundary⁹³. Even this, however, would indicate an earlier date for the common ancestry of the Fayum simians. Identification of five new fossil simian genera very low down in the deposits (largely based on fragmentary specimens, but including the relatively well preserved skull of *Catopithecus*⁹⁴⁻⁹⁷) have reinforced the effects of this redating to indicate that simians must have emerged by the late Eocene, if not earlier. The new discoveries have raised the total number of simian genera reported from the Fayum to 11 (*Aegyptopithecus*, *Apidium*, *Arsinoea*, *Catopithecus*, *Oligopithecus*, *Parapithecus*, *Plesiopithecus*, *Propliopithecus*, *Proteopithecus*, *Qatrania*, *Serapia*), indicating that considerable diversification had taken place at an early stage. One of the new early forms (*Plesiopithecus*) may have lacked incisors in the lower jaw. If the identification of *Plesiopithecus* as a simian on the basis of a single lower jaw fragment is correct, considerable diversity had apparently already developed following divergence from ancestral simians. On the basis of isolated teeth, simians have also now been reported from earliest Oligocene deposits in Oman (including *Oligopithecus*)⁹⁸ and from late Eocene deposits in Algeria (*Biretia*)⁹⁹.

Indications of an earlier date for the Fayum simians may have been superseded by the recovery of dental specimens representing an apparent simian, *Algeripithecus minutus*, from the Algerian site of Glib Zegdou, which is of middle or even early Eocene age¹⁰⁰. The find is unfortunately confined to three cheek teeth, but the upper and lower molars exhibit a number of simian-like features. In particular, certain upper molar features are shared with *Aegyptopithecus* from the Fayum deposits. If these features are indicative of a sister-group relationship between *Algeripithecus* and *Aegyptopithecus*, as proposed¹¹, then the origin of known simians must be pushed back to the lower Eocene or even into the Palaeocene. Against this background,

the late Eocene *Amphipithecus*¹⁰¹ and *Pondaungia*¹⁰² from Burma clearly now deserve serious re-examination. Both of these fossil genera are of uncertain status but could well be linked to simian origins⁷⁹. Overall, a date of at least 55 Myr instead of just over 35 Myr for the origin of simians now seems possible. It is, in fact, predictable that the simians originated far earlier than hitherto accepted by most authors (See Box 1) and an even earlier date for simian ancestry cannot be ruled out. Nevertheless, although it now seems quite probable that the simian primates emerged much earlier than previously believed, their origins remain obscure (Box 2).

Future developments

It is clear that there are still many uncertainties regarding the course and timing of early primate evolution. There are, however, several indications that conclusions drawn from an exploratory assessment of the sampling density of the primate fossil record (Box 1) are correct in principle. Our current sample of the primate fossil record is very limited and it follows from this that times of divergence within the primate tree may have been commonly underestimated by a considerable margin. Against this background, it would not be surprising if the time of origin of simian primates were to be pushed back into the Palaeocene, in which case direct migration of simians between Africa and South America¹⁰³ becomes far more likely. Limited sampling of the fossil record, combined with the fragmentary nature of most of the fossils concerned, also explains why interpretations of primate evolution have been subject to repeated, often extensive revision. In the face of major gaps in the fossil record, far-reaching interpretation of fragmentary fossil remains can easily lead to misinterpretation of phylogenetic relationships. As a general rule, there has been a tendency to overlook the significance of convergent evolution in the interpretation of fossil primates. If only a limited number of characters can be investigated, it is extremely difficult to test for possible convergence. Greater caution in the interpretation of fragmentary fossil specimens combined with a healthy respect for convergent evolution might help to reduce the level of misinterpretation in the future.

Recognition of the implications of the low sampling level of the primate fossil record in fact raises a new possibility with respect to the evolutionary relationships of the early Tertiary adapids and omomyids. At present, modern strepsirhines, tarsiers and simians have been traced back to adapids or omomyids because these are the only known fossil forms. Rather than seeking to establish direct links with modern primate groups, we should perhaps consider the possibility that the adapids and omomyids together are the outcome of an entirely separate radiation of early primates of modern aspect in the Northern Hemisphere. The divergence between adapids and omomyids could have evolved in parallel to a divergence between strepsirhine and haplorhine primates in the Southern Hemisphere, such that adapids and omomyids are in fact more closely related to one another than to any modern primates. Such an interpretation, which (like all of the potential solutions reviewed above) requires acceptance of a certain amount of convergent evolution, could explain certain features that clash with reconstructions linking adapids to modern strepsirhines and omomyids to modern haplorhines or with alternative hypotheses. It is, for instance, unusual that the adapids were typically larger than omomyids, whereas the reverse is the case for the comparison of modern strepsirhines with haplorhines¹⁰⁴. Adapids were presumably predominantly diurnal and yet supposedly gave rise to the primary nocturnal strepsirhines, whereas the omomyids were presumably predominantly nocturnal and yet supposedly gave rise to the primarily diurnal haplorhines¹⁴. As with many other aspects of early primate evolution, more precise answers depend on a continuing, vital effort by field palaeontologists, who will hopefully uncover many more unknown extinct primate species.

Robert D. Martin is at the Anthropologisches Institut und Museum, Universität Zürich-Irchel, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland.

1. Szalay, F. S. & Delson, E. *Evolutionary History of the Primates* (Academic, New York, 1979).
2. Beard, K. C. *Nature* **345**, 340–341 (1990).
3. Kay, R. F., Thorington, R. W. & Houde, P. *Nature* **345**, 342–344 (1990).
4. Martin, R. D. *Nature* **345**, 291–292 (1990).
5. Beard, K. C. *Postcranial Anatomy, Locomotor Adaptations, and Paleoeecology of Early Cenozoic Plasiadapidae, Paromomyidae and Micromomyidae* (Eutheria, Dermoptera) (Johns Hopkins University, Baltimore, 1989).
6. Beard, K. C. in *Origine(s) de la Bipedie chez les Homínides* (eds Coppens, Y. & Senut, B.) 79–87 (Éditions du CNRS, Paris, 1991).
7. Beard, K. C. *J. vert. Paleont.* **9**, 13A (1989).
8. Kay, R. F., Thewissen, J. G. M. & Yoder, A. D. *Am. J. phys. Anthropol.* **89**, 477–498 (1992).
9. Gunnell, G. F. *Univ. Mich. Pap. Paleont.* **27**, 1–157 (1989).
10. Hoffstetter, R. C. *r. Acad. Sci. Paris ser. II* **302**, 43–45 (1986).
11. Gingerich, P. D. *Univ. Mich. Pap. Paleont.* **15**, 1–140 (1976).
12. MacPhee, R. D. E., Cartmill, M. & Gingerich, P. D. *Nature* **301**, 509–511 (1983).
13. Beard, K. C. in *Mammal Phylogeny Vol. 2. Placental* (eds Szalay, F. S., Novacek, M. J. & McKenna, M. C.) 129–150 (Springer, New York, in the press).
14. Martin, R. D. *Primate Origins and Evolution: A Phylogenetic Reconstruction* (Chapman Hall/Princeton Univ. Press, London/New Jersey, 1990).
15. Krause, D. W. *J. hum. Evol.* **21**, 177–188 (1991).
16. Hunt, R. B. & Korth, W. W. *J. Morph.* **164**, 167–211 (1980).
17. Rose, K. D. & Simons, E. L. *Contrib. Mus. Paleont. Univ. Michigan* **24**, 221–236 (1977).
18. MacPhee, R. D. E., Cartmill, M. & Rose, K. D. *J. vert. Paleont.* **9**, 329–349 (1989).
19. Ducrocq, S. et al. *Palaeontology* **35**, 373–380 (1982).
20. Russell, D. E. et al. *Paleovertebrata Mémoire Extraordinaire* 1–77 (1982).
21. Pirlot, P. & Kamiya, T. *Can. J. Zool.* **60**, 565–572 (1982).
22. Lekagul, B. & McNeely, J. A. *Mammals of Thailand 1–758* (Sahakarnbhat, Bangkok, 1977).
23. Gregory, W. K. *Bull. Am. Mus. nat. Hist.* **27**, 1–524 (1910).
24. Butler, P. M. in *Studies in Vertebrate Evolution* (eds Joysey, K. A. & Kemp, T. S.) 253–265 (Oliver and Boyd, Edinburgh, 1972).
25. McKenna, M. C. in *Phylogeny of the Primates* (eds Luckett, W. P. & Szalay, F. S.) 21–46 (Plenum, New York, 1975).
26. Szalay, F. S. & Drawhorn, G. in *Comparative Biology and Evolutionary Relationships of Tree Shrews* (ed. Luckett, W. P.) 133–169 (Plenum, New York, 1980).
27. Cronin, J. E. & Sarich, V. M. in *Comparative Biology and Evolutionary Relationships of the Tree Shrews* (ed. Luckett, W. P.) 293–312 (Plenum, New York, 1980).
28. Shoshani, J. *Molec. Biol. Evol.* **3**, 222–242 (1986).
29. Novacek, M. J. & Wyss, A. R. *Cladistics* **2**, 257–287 (1988).
30. Novacek, M. J., Wyss, A. R. & McKenna, M. C. in *The Phylogeny and Classification of the Tetrapods* (ed. Benton, M. J.) 99–116 (Clarendon, Oxford, 1988).
31. Novacek, M. J. *Nature* **356**, 121–125 (1992).
32. Goodman, M., Czelusniak, J. & Beebe, J. E. *Cladistics* **1**, 171–185 (1985).
33. Miyamoto, M. M. & Goodman, M. *Syst. Zool.* **35**, 230–240 (1986).
34. Pettigrew, J. D. *Science* **231**, 1304–1306 (1986).
35. Pettigrew, J. D. et al. *Phil. Trans. R. Soc. B325*, 489–559 (1989).
36. Pettigrew, J. D. *Am. J. phys. Anthropol.* **18**, suppl. 3–54 (1991).
37. Wible, J. R. & Novacek, M. J. *Am. Mus. Novitates* **2911**, 1–19 (1988).
38. Baker, R. J., Novacek, M. J. & Simons, N. B. *Syst. Zool.* **40**, 216–231 (1991).
39. Thiele, A., Vogelsang, M. & Hoffmann, K.-P. *J. comp. Neurol.* **314**, 671–683 (1991).
40. Bailey, W. J., Slightom, J. L. & Goodman, M. *Science* **256**, 86–89 (1992).
41. Adkins, R. M. & Honeycut, R. L. *Proc. natn. Acad. Sci. U.S.A.* **88**, 10317–10321 (1991).
42. Wyss, A. R., Novacek, M. J. & McKenna, M. C. *Molec. Biol. Evol.* **4**, 99–116 (1987).
43. Krebs, B. *Berliner geowiss. Abh.* **133**, 1–121 (1991).
44. Gregory, W. K. *Mem. Am. Mus. nat. Hist.* **3**, 49–243 (1920).
45. Gazin, C. L. *Smithson. Misc. Coll.* **136**, 1–112 (1958).
46. Gingerich, P. D. in *Evolutionary Biology of the New World Monkeys and Continental Drift* (eds Ciochon, R. L. & Chiarelli, A. B.) 123–138 (Plenum, New York, 1980).
47. Gebo, D. L., Dagosto, M. & Rose, K. D. *Am. J. phys. Anthropol.* **86**, 51–72 (1991).
48. Beard, K. C., Dagosto, M., Gebo, D. L. & Godinot, M. *Nature* **331**, 712–714 (1988).
49. Godinot, M. & Beard, K. C. *Hum. Evol.* **6**, 307–354 (1991).
50. Rosenberger, A. L., Strasser, E. & Delson, E. *Folia Primatol.* **44**, 15–39 (1985).
51. Stehlin, H. G. *Abh. schweiz. pal. Ges.* **38**, 1165–1298 (1912).
52. Lanéque, L. C. *r. Acad. Sci. Paris ser. II* **314**, 1387–1393 (1992).
53. Lanéque, L. *Hum. Evol.* **7**, 1–16 (1992).
54. Gingerich, P. D. *Folia Primatol.* **28**, 60–80 (1977).
55. Gingerich, P. D. *Géobios Mém. spéc.* **1**, 165–182 (1977).
56. Godinot, M. C. *r. Acad. Sci. Paris ser. II* **314**, 237–242 (1992).
57. Gingerich, P. D. *Am. J. phys. Anthropol.* **56**, 217–234 (1981).
58. Gingerich, P. D. & Martin, R. D. *Am. J. phys. Anthropol.* **56**, 235–257 (1981).
59. Krishtalka, L., Stucky, R. K. & Beard, K. C. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5223–5226 (1990).
60. Martin, R. D. *Phil. Trans. R. Soc. B264*, 295–352 (1972).
61. Martin, R. D. in *The Study of Prosimian Behavior* (eds Doyle, G. A. & Martin, R. D.) 45–77 (Academic, New York, 1979).
62. Dagosto, M. *Folia Primatol.* **41**, 49–101 (1983).
63. Dagosto, M. & Terranova, C. J. *Int. J. Primatol.* **13**, 307–344 (1992).
64. von Koenigswald, W. *Paläont. Z.* **53**, 63–76 (1979).
65. von Koenigswald, W. *Carolinea* **42**, 145–148 (1985).
66. Franzen, J. L. *Cour. Forsch.-Inst. Senckenberg* **91**, 151–187 (1987).
67. Franzen, J. L. *Cour. Forsch.-inst. Senckenberg* **107**, 263–273 (1988).
68. Franzen, J. L. & Frey, E. *Abstract in International Conference Monument Grube Messel* (Darmstadt, 1991).
69. Thalmann, U., Haubold, H. & Martin, R. D. *Paleovertebrata* **19**, 115–130 (1989).
70. Moya Sola, S. & Köhler, M. *Folia Primatol.* (in the press).
71. Gingerich, P. D. & Schoeninger, M. J. *J. hum. Evol.* **6**, 483–505 (1977).
72. Franzen, J. L. & Haubold, H. *Mod. Geol.* **10**, 159–170 (1986).
73. Suteethorn, V., Buffetaut, E., Helmcke-Ingavat, R., Jaeger, J.-J. & Jongkanjasontorn, Y. *N. Jb. Geol. Paläont. Mh.* **9**, 563–570 (1988).
74. Rosenberger, A. L. *Folia Primatol.* **45**, 179–194 (1985).
75. Rosenberger, A. L. & Szalay, F. S. in *Evolutionary Biology of the New World Monkeys and Continental Drift* (eds Ciochon, R. C. & Chiarelli, A. B.) 139–157 (Plenum, New York, 1980).
76. Schmid, P. in *Primate Evolutionary Biology* (eds Chiarelli, A. B. & Corruccini, R. S.) 6–13 (Springer, Berlin, 1981).
77. Schmid, P. *Die Systematische Revision der europäischen Microchoeridae Lydekker, 1887 (Omomyiformes, Primates)* (University of Zürich, 1982).

78. MacPhee, R. D. E. & Cartmill, M. in *Comparative Primate Biology* (eds Swindler, D. R. & Erwin, J.) 219–275 (Liss, New York, 1986).
79. Gingerich, P. D. *Yb. phys. Anthropol.* **27**, 57–72 (1984).
80. Aiello, L. C. in *Major Topics in Primate and Human Evolution* (eds Wood, B. A., Martin, L. B. & Andrews, P. J.) 47–65 (Cambridge Univ. Press, Cambridge, 1986).
81. Simons, E. L. & Rasmussen, D. T. *Am. J. phys. Anthropol.* **79**, 1–23 (1989).
82. Schmid, P. *Folia Primatol.* **40**, 1–10 (1983).
83. Cartmill, M. in *Evolutionary Biology of the New World Monkeys and Continental Drift* (eds Ciochon, R. L. & Chiarelli, A. B.) 243–274 (Plenum, New York, 1980).
84. Rosenberger, A. L. & Dagosto, M. in *Topics in Primatology Vol. 3 Evolutionary Biology, Reproductive Endocrinology, and Virology* (eds Matano, S., Tuttle, R. H., Ishida, H. & Goodman, M.) 37–51 (Univ. of Tokyo Press, Tokyo, 1992).
85. Schmid, P. *Folia Primatol.* **31**, 301–311 (1979).
86. Simons, E. L. *Primate Evolution: An Introduction to Man's Place in Nature* (Macmillan, New York, 1972).
87. Beard, K. C., Krishtalka, L. & Stucky, R. K. *Nature* **349**, 64–67 (1991).
88. Martin, R. D. *Nature* **349**, 19–20 (1991).
89. Jouffroy, F.-K., Berge, C. & Niemitz, C. in *Biology of Tarsiers* (eds Niemitz, C.) 167–190 (Gustav Fischer, Stuttgart, 1984).
90. Sigé, B., Jaeger, J.-J., Sudre, J. & Vianey-Liaud, M. *Palaeontographica Abt. A* **214**, 31–56 (1990).
91. Fleagle, J. G., Bown, T. M., Obradovitch, J. D. & Simons, E. L. *Science* **234**, 1247–1249 (1986).
92. Van Couvering, J. A. & Harris, J. A. *J. hum. Evol.* **21**, 241–260 (1991).
93. Kappelman, J. *J. hum. Evol.* **6**, 495–503 (1992).
94. Simons, E. L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9956–9960 (1989).
95. Simons, E. L. *Science* **247**, 1567–1569 (1990).
96. Rasmussen, D. T. & Simons, E. L. *Int. J. Primatol.* **13**, 477–508 (1992).
97. Simons, E. L. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10743–10747 (1992).
98. Thomas, H., Roger, J., Sen, S. & Al-Sulaimani, Z. C. r. *Acad. Sci. Paris, sér. II* **306**, 823–829 (1988).
99. de Bonis, L., Jaeger, J.-J., Coiffait, B. & Coiffait, P.-E. C. r. *Acad. Sci. Paris sér. II* **306**, 929–934 (1988).
100. Godinot, M. & Mahboubi, M. *Nature* **357**, 324–326 (1992).
101. Ciochon, R. L., Savage, D. E., Thaw, T. & Maw, B. *Science* **229**, 756–759 (1985).
102. Maw, B., Ciochon, R. L. & Savage, D. E. *Nature* **282**, 65–67 (1979).
103. Hoffstetter, R. in *Evolutionary Biology of the New World Monkeys and Continental Drift* (eds Ciochon, R. C. & Chiarelli, A. B.) 103–188 (Plenum Press, New York, 1980).
104. Fleagle, J. G. *Paleobiology* **4**, 67–76 (1978).
105. Martin, R. D. in *Major Topics in Primate and Human Evolution* (eds Wood, B. A., Martin, L. B. & Andrews, P. J.) 1–31 (Cambridge Univ. Press, Cambridge, 1986).
106. Wood Jones, F. *The Ancestry of Man* 1–35 (Gillies, Brisbane, 1923).
107. Wortman, J. L. *Am. J. Sci. ser. 4* **17**, 203–214 (1904).
108. Gidley, J. W. *Proc. U. S. Nat. Mus.* **63**, 1–38 (1923).
109. Hoffstetter, R. *J. hum. Evol.* **3**, 327–350 (1974).
110. Hershkovitz, P. *Folia Primatol.* **21**, 1–35 (1974).
111. Rasmussen, D. T. *J. hum. Evol.* **15**, 1–12 (1986).
112. Rasmussen, D. T. *Am. J. Primatol.* **22**, 263–277 (1990).
113. Rasmussen, D. T. & Simons, E. L. *Folia Primatol.* **51**, 182–208 (1989).
114. Rasmussen, D. T. *Int. J. Primatol.* **11**, 439–469 (1990).
115. Beecher, R. M. *Am. J. phys. Anthropol.* **50**, 418 (1979).
116. Beecher, R. M. *Int. J. Primatol.* **4**, 99–112 (1983).
117. Ginsburg, L. & Mein, P. C. r. *Acad. Sci. Paris sér. II* **304**, 1213–1215 (1986).

Global climate change and terrestrial net primary production

Jerry M. Melillo*, A. David McGuire*, David W. Kicklighter*, Berrien Moore III†, Charles J. Vorosmarty† & Annette L. Schloss†

* The Ecosystems Center, Marine Biological Laboratory, Woods Hole, Massachusetts 02543, USA

† Complex Systems Research Center, Institute for the Study of Earth, Oceans and Space, University of New Hampshire, Durham, New Hampshire 03824, USA

A process-based model was used to estimate global patterns of net primary production and soil nitrogen cycling for contemporary climate conditions and current atmospheric CO₂ concentration. Over half of the global annual net primary production was estimated to occur in the tropics, with most of the production attributable to tropical evergreen forest. The effects of CO₂ doubling and associated climate changes were also explored. The responses in tropical and dry temperate ecosystems were dominated by CO₂, but those in northern and moist temperate ecosystems reflected the effects of temperature on nitrogen availability.

THE atmospheric concentrations of the major long-lived greenhouse gases continue to increase because of human activity¹. Most climate models predict that the buildup of these gases is likely to lead to surface air temperature rises of 1.5 °C to 4.5 °C and changes in precipitation and cloud patterns over the next century². Climate changes of this magnitude are expected to affect the net primary production (NPP) of the world's land ecosystems³. Annual NPP is the net amount of carbon captured by land plants through photosynthesis each year. It is of fundamental importance to humans because the largest portion of our food supply is from productivity of plant life on land, as is wood for construction and fuel. Because climate changes are predicted to vary from place to place², estimating the response of NPP will require the use of models that can make geographically referenced predictions. Both regression- and process-based models are available to assess the response of terrestrial NPP with some degree of geographic specificity.

Regression-based models use empirically derived relationships between climate and NPP to make predictions⁴. Although these models can presently be extrapolated for all land ecosystems^{5–7}, their use for predicting NPP responses is limited because the regressions may not be appropriate for climatic conditions that are novel to terrestrial ecosystems⁸. Unlike regression-based models, process-based models describe how important ecosystem processes such as photosynthesis, respiration, decomposition and nutrient cycling interact to affect NPP. Therefore they have the potential for accurately describing how these processes

will interact in future climates⁸. Although process-based models have been used in regional studies to evaluate responses of NPP to climate change in a geographically referenced manner^{8–12}, none has been applied globally.

Here we report the results of a study in which we use a process-based terrestrial ecosystem model (TEM) to estimate global patterns of NPP for contemporary climate conditions and current atmospheric CO₂. In addition, we use the output from global climate models, known as general circulation models (GCMs), to estimate the potential effects of a CO₂ doubling and associated climate changes on NPP for the world's land ecosystems. We have restricted our analysis to evaluate the equilibrium response of NPP to changes in CO₂ and climate for the vegetation distribution that is appropriate to contemporary climate. We do not consider how the response of vegetation distribution to climate change^{13–17} affects NPP.

The terrestrial ecosystem model

The TEM (Fig. 1) is a process-based ecosystem simulation model that uses spatially referenced information on climate, elevation, soils, vegetation and water availability to make monthly estimates of important carbon and nitrogen fluxes and pool sizes^{8,12,18}. Because we use TEM to make equilibrium predictions, its estimates of carbon and nitrogen dynamics apply only to mature, undisturbed vegetation; they do not include the effects of land use.

TABLE 1 Values of vegetation-specific parameters used in the TEM

Vegetation type	P_{cn}^*	$C_{max}^\dagger$	$K_r^\dagger$	$K_d^\dagger$	$KFALL^\dagger$	$N_{max}^\dagger$	$N_{up}^\dagger$	$NFALL^*$	$T_{min}^\dagger$	$T_{opt}^\dagger$	$T_{max}^\dagger$	$V_{CN}^\dagger$
Polar desert/alpine tundra	32.50	591.80	0.038900	0.001048	0.012037	1.51000	0.706693	0.006325	-1.0	15.0	36.0	69.23
Wet/moist tundra	30.00	955.50	0.038900	0.000645	0.013333	3.48000	1.533380	0.004410	-1.0	15.0	36.0	50.00
Boreal woodland	45.24	842.00	0.011080	0.000924	0.005871	0.85500	0.205103	0.004692	-1.0	15.0	36.0	91.67
Boreal forest	52.38	676.20	0.002185	0.001396	0.002037	0.70900	0.081884	0.007950	-1.0	15.0	36.0	375.00
Temperate coniferous forest	89.17	1,013.90	0.001316	0.001219	0.001025	0.50000	0.034516	0.004667	-1.0	18.0	39.0	580.00
Desert	30.56	492.00	0.006180	0.000973	0.016975	0.53850	0.142026	0.011540	1.0	31.0	51.0	27.69
Arid shrubland	30.56	492.00	0.006180	0.000973	0.016975	0.53850	0.142026	0.011540	1.0	31.0	51.0	27.69
Short grassland	54.29	779.20	0.017150	0.004323	0.052910	0.45950	0.247182	0.033020	0.0	27.0	45.0	35.80
Tall grassland	69.55	977.35	0.016775	0.001008	0.054487	0.29080	0.138580	0.074720	0.0	27.0	45.0	108.33
Temperate savanna	66.44	1,083.70	0.006975	0.004723	0.018254	0.47020	0.157422	0.029783	-1.0	24.0	46.0	131.25
Temperate deciduous forest	65.00	1,207.90	0.001465	0.002303	0.003483	0.75300	0.073770	0.018090	-1.0	20.0	43.0	426.05
Temperate mixed forest	74.97	1,125.90	0.002255	0.002422	0.003660	0.49862	0.076415	0.015400	-1.0	19.0	41.0	422.29
Temperate broadleaf evergreen forest	90.63	780.75	0.001833	0.002110	0.004028	0.47560	0.073452	0.011905	0.0	25.0	49.0	357.14
Mediterranean shrubland	39.29	614.64	0.003138	0.001841	0.010659	1.10000	0.118475	0.013550	-1.0	25.0	49.0	47.78
Tropical savanna	33.23	1,897.75	0.005481	0.000976	0.005747	0.64650	0.067100	0.007166	1.0	30.0	49.0	41.43
Xeromorphic forest	39.29	614.64	0.003138	0.001841	0.010659	1.10000	0.118475	0.013550	-1.0	25.0	49.0	47.78
Tropical deciduous forest	32.66	1,426.40	0.002657	0.002242	0.005140	3.41400	0.197320	0.010236	0.0	27.0	48.0	66.76
Tropical evergreen forest	43.75	1,034.75	0.001346	0.001250	0.003889	2.83000	0.111000	0.006690	2.0	28.0	48.0	75.00

See Table 2 in ref. 12 for the parameter values of the leaf phenology model, Table 4 in ref. 12 for values of the soil-specific parameters, and Table A2 in ref. 18 for values of the constant parameters k_i , k_{n1} and k_{n2} .

* See ref. 12 for parameter definition.

† See ref. 18 for parameter definition.

For each monthly time step in a model run, NPP is calculated as the difference between gross primary productivity (GPP) and plant respiration (R_A). The calculation of R_A considers both maintenance respiration^{8,12} and construction respiration¹⁸. The flux GPP considers the effects of several factors and is calculated at each time step as follows:

$$GPP = C_{max} f(PAR) f(LEAF) f(T) f(CO_2, H_2O) f(NA)$$

where C_{max} is the maximum rate of C assimilation, PAR is photosynthetically active radiation, LEAF is leaf area relative to maximum annual leaf area, T is temperature, CO_2 is atmospheric carbon dioxide, H_2O is water availability, and NA is nitrogen availability. All of the functions in the GPP equation, as well as other mathematical expressions in the model, are well documented in previous work^{8,12,18}. Here we review the descriptions of $f(CO_2, H_2O)$ and $f(NA)$ because of their importance in affecting the capacity of the vegetation to incorporate elevated CO_2 into production.

The function $f(CO_2, H_2O)$ is described by the hyperbolic relationship^{8,18}:

$$f(CO_2, H_2O) = C_i / (k_c + C_i)$$

where C_i is the concentration of CO_2 within leaves of the canopy and k_c is the half-saturation constant for CO_2 uptake by plants.

FIG. 1 The terrestrial ecosystem model (TEM). Carbon enters the vegetation pool (C_V) as gross primary productivity (GPP) and transfers to the atmosphere as plant respiration (R_A) or to the soil pool (C_S) as litter production (L_C); it leaves the soil as heterotrophic respiration (R_H). Nitrogen inputs from outside the ecosystem (NINPUT) enter the inorganic N pool (N_{AV}); losses leave this pool as the flux NLOST. Nitrogen in the vegetation occurs either in the structural pool (N_{VS}) or the labile pool (N_{VL}). Structural N in vegetation is constructed from N that is derived from either the labile pool as the flux NMOBIL or from soil inorganic N pool as the flux NUPTAKE_S. The labile pool is replenished from N that is resorbed from senescing tissue (NRESORB), N that is allocated for storage (NMOBIL), or N in uptake that does not enter directly into tissue construction (NUPTAKE_L). Nitrogen is transferred from vegetation to the soil organic pool (N_S) as the flux L_N . Net N mineralization (NETNMIN) accounts for N exchanged between the organic and inorganic N pools of the soil.

The variable C_i is the product of ambient CO_2 and relative canopy conductance to CO_2 , a variable which increases from 0 to 1 with increasing water availability^{12,18}. The parameter k_c has been chosen to increase $f(CO_2, H_2O)$ by 37% for a doubling of atmospheric CO_2 from 340 parts per million by volume (p.p.m.v.) to 680 p.p.m.v. with canopy conductance equal to 1 (refs 8, 12). Among studies that have provided adequate water and nutrients to experimental plants, the range in the response of plant growth to doubled CO_2 is between 24% and 50% (refs. 19, 20).

The function $f(NA)$ models the limiting effects of plant nitrogen status on GPP^{8,12}. It constrains C uptake when N supply, defined as the combination of N uptake and vegetation labile N, limits production. Information on the C to N ratio of production (P_{cn}), a quantity commonly measured in ecosystem studies, is used to determine when N supply limits production. This implementation assumes that nitrogen use efficiency, defined as the ratio of NPP to N in new production, is conservative within a vegetation type.

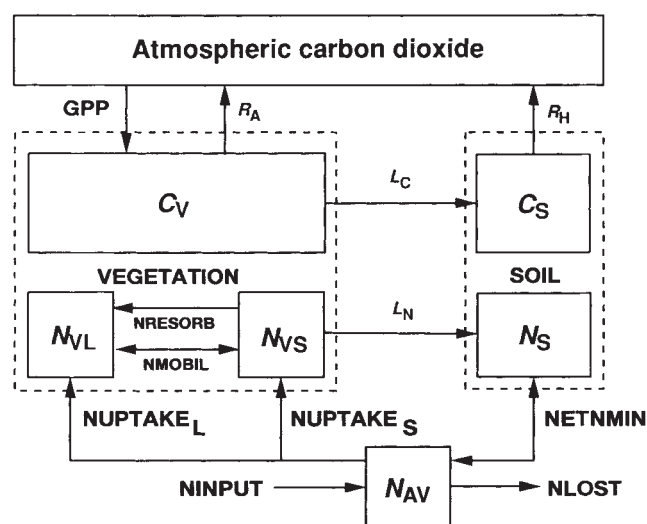


TABLE 2 Estimates by the TEM of annual NPP and nitrogen uptake for potential vegetation in the terrestrial biosphere at an atmospheric concentration of 355 p.p.m.v. CO₂

Vegetation type	Area (10 ⁶ km ²)	Cells	Total NPP (10 ¹⁵ g C yr ⁻¹)	Mean NPP (g C m ⁻² yr ⁻¹)	Max NPP (g C m ⁻² yr ⁻¹)	Min NPP	Total N uptake (10 ¹² g N yr ⁻¹)	Mean N Uptake (g N m ⁻² yr ⁻¹)
Polar desert/alpine								
tundra	5.0	3,147	0.4	87	216	0	3	0.7
Wet/moist tundra	4.7	3,788	0.6	120	423	34	4	0.8
Boreal woodland	6.3	4,414	1.1	173	420	89	9	1.5
Boreal forest	12.2	7,406	2.9	238	434	124	31	2.5
Temperate coniferous forest	2.4	1,081	1.1	465	704	208	9	3.7
Desert	11.5	4,145	0.6	53	370	0	15	1.3
Arid shrubland	14.5	5,708	1.9	129	454	6	46	3.2
Short grassland	4.7	2,050	1.0	214	438	72	17	3.7
Tall grassland	3.6	1,557	1.2	335	756	136	16	4.4
Temperate savanna	6.8	2,886	2.3	342	785	68	29	4.3
Temperate mixed forest	5.1	2,250	3.4	669	1,066	231	37	7.3
Temperate deciduous forest	3.5	1,614	2.2	620	978	81	27	7.6
Temperate broadleaf evergreen forest	3.2	1,205	2.4	741	1,001	322	20	6.2
Mediterranean shrubland	1.4	554	0.5	343	634	32	12	8.7
Tropical savanna	13.7	4,624	5.4	393	786	88	162	11.8
Xeromorphic forest	6.8	2,357	3.1	461	992	0	79	11.7
Tropical deciduous forest	4.6	1,577	4.0	871	1,398	323	121	26.2
Tropical evergreen forest	17.4	5,727	19.1	1,098	1,422	407	436	25.1
Total	127.3	5,6090	53.2	418	1,422	0	1,073	8.4

Ecosystem-based estimates may not sum to totals because of the effects of rounding in reporting those estimates.

The data sets used to drive TEM are gridded at a resolution of 0.5° latitude by 0.5° longitude. The sources for the global data sets on climate (air temperature, precipitation and cloudiness), elevation and soil texture are described elsewhere¹⁸; the climate data represent long-term averages. The data set of global potential vegetation (Fig. 2) was constructed from a number of sources^{21–31}. Hydrological inputs for TEM were determined with a water balance model³² that uses the climate, elevation, soils and vegetation data.

The application of TEM to a grid cell requires the use of monthly climatic and hydrological data and the soil- and vegetation-specific parameters appropriate to the grid cell. Although many of the vegetation-specific parameters in the model (Table 1) are defined from published information^{12,18}, some are determined by calibrating the model to the fluxes and pool sizes of an intensively studied field site. Most of the data used to calibrate the model for the 18 vegetation types considered in this study are documented elsewhere¹². We do not make estimates for grid cells defined as ice, open water, or wetland ecosystems, so that our global extrapolation of TEM requires application of the model to 56,090 grid cells in the terrestrial biosphere. Because wetlands are represented by only 1,466 grid cells, we do not expect their exclusion to affect global estimates of terrestrial NPP substantially.

Global extrapolation for contemporary climate

To estimate carbon and nitrogen dynamics of potential vegetation for 'contemporary' conditions, we applied TEM globally at 355 p.p.m.v. CO₂ using the long-term climate data. Under these conditions, TEM estimates the global annual NPP for potential vegetation to be 53.2 Pg C (10¹⁵ g C), or 418 g C m⁻² yr⁻¹ (Table 2). Our process-based estimate is similar to many of the estimates that have appeared in the literature^{5–7,33–42} (mean: 53.1 Pg C; *N* = 13; range: 40.5 Pg C to 78.0 Pg C; s.d. 9.3 Pg C). Most of the estimates have been calculated by multiplying the mean NPP for an ecosystem, as determined from a literature survey of field studies, by the area of the ecosystem, and then summing across ecosystems^{33–41}. A few of the estimates have been determined by regression approaches

involving climate variables^{5–7} and some consider the effects of land use^{6–7,42}.

Over half of the global annual NPP occurs in the tropics between the latitudes of 22.5° S and 22.5° N (29.6 Pg C yr⁻¹; Fig. 2). Most of this productivity is attributable to tropical evergreen forest which accounts for 35.9% of the net exchange of CO₂ between terrestrial vegetation and the atmosphere, although it covers only 13.7% of the terrestrial land surface (Table 2). The least productive vegetation types include polar desert, tundra, and desert, which collectively account for 3.0% of terrestrial NPP and cover 16.7% of the terrestrial land area (Table 2). Mean NPP estimates for vegetation types range from 53 g C m⁻² yr⁻¹ for desert to 1,098 g C m⁻² yr⁻¹ for tropical evergreen forest (Table 2). Estimates for individual grid cells range from 0 g C m⁻² yr⁻¹ to 1,422 g C m⁻² yr⁻¹ (Table 2). The variability of NPP estimates by TEM, which reflects spatial heterogeneity in climate and soils, has been evaluated in previous applications of the model^{12,18}. The NPP predictions of this study generally compare well with the field measurements used to construct several regression-based global models of NPP⁵ (Fig. 3).

Annual nitrogen uptake by global potential vegetation under conditions of contemporary climate and CO₂ is estimated to be 1,073 Tg N (10¹² g N), or 8.4 g N m⁻² yr⁻¹ (Table 2). Assuming a global nitrogen-use efficiency (NPP/nitrogen uptake) of 50, others have estimated the annual nitrogen cycling in the terrestrial biosphere to be 1,200 Tg N (ref. 43) and 1,400 Tg N (ref. 44). The global nitrogen-use efficiency in this study is also estimated to be 50 (53.2 Pg C yr⁻¹/1,073 Tg N yr⁻¹). Mean estimates of nitrogen uptake for ecosystems range from 0.7 g N m⁻² yr⁻¹ for polar desert to 26.2 g N m⁻² yr⁻¹ in tropical deciduous forest (Table 2). The accuracy of nitrogen cycling estimates by TEM has been evaluated for several ecosystems in a previous application of the model¹².

Future climate scenarios for use with TEM

We obtained the output of four GCMs from the National Center for Atmospheric Research⁴⁵. The simulations estimate equilibrium climates that correspond to a doubling of the atmospheric

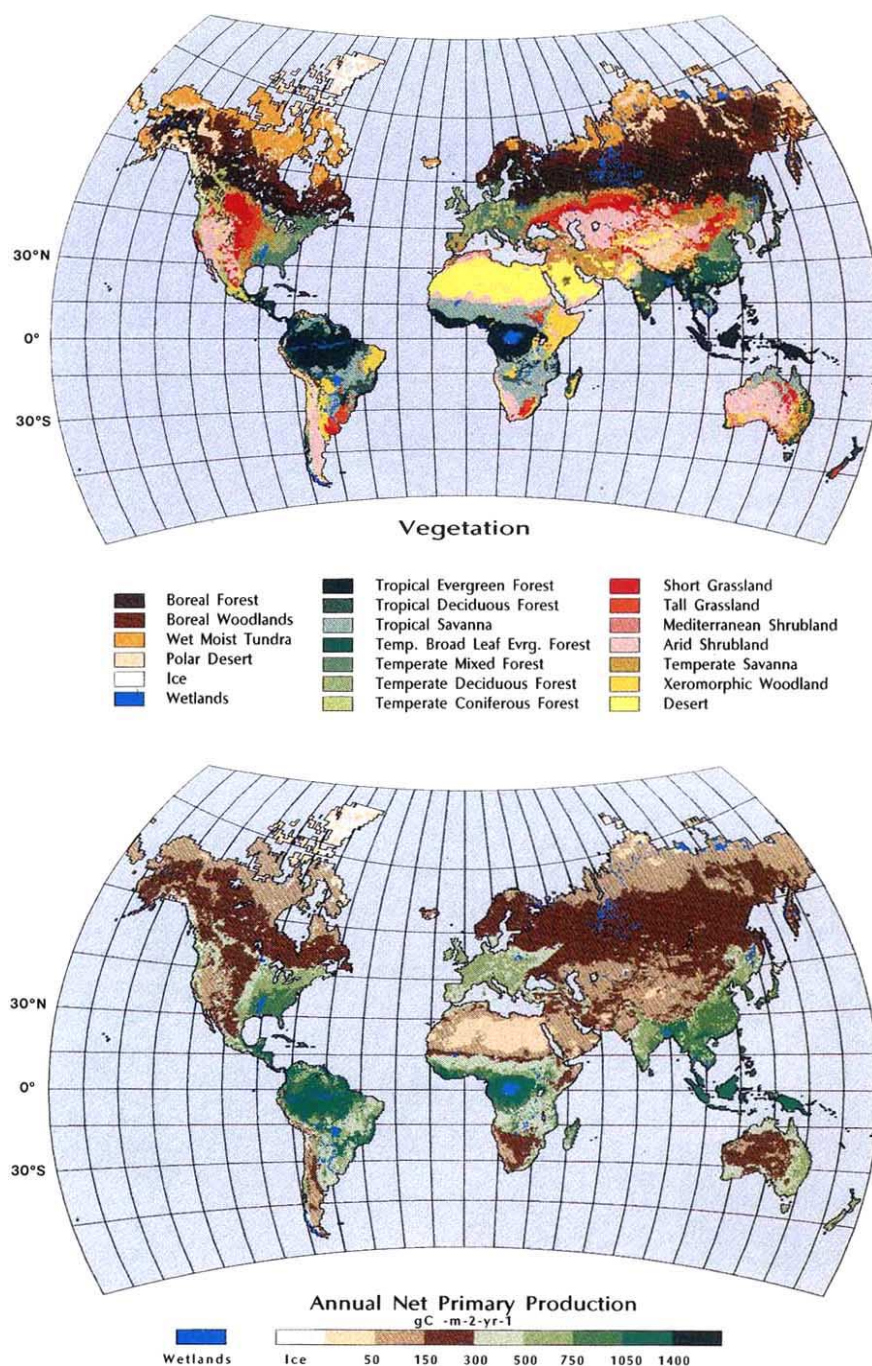


FIG. 2 Potential natural vegetation (top) used to estimate annual NPP (bottom) for the global terrestrial biosphere.

CO₂ concentration and include: the Goddard Institute of Space Studies GCM (GISS); the Oregon State University GCM (OSU); and two GCM simulations from the Geophysical Fluid Dynamics Laboratory (GFDL1 and GFDLQ). Among the GCMs, mean global temperature increases between 2.8 °C and 4.2 °C, global precipitation increases between 7.8% and 11.0%, and global cloudiness decreases between 0.4% and 3.4%.

Because we were interested in the implications of climate change for NPP, we generated 'GCM climates' for TEM by using the output variables of surface air temperature, precipitation, and total cloud cover for the current and 2×CO₂ simulations of each GCM to modify the contemporary climate data for TEM. For the heuristic purpose of examining implications at the spatial scale of TEM predictions, we organized each of the output variables of each GCM simulation at the 0.5°×0.5° resolution with a spherical interpolation procedure⁴⁶. Next, similar to the method used in a study of the potential effects of climate change on US agriculture⁴⁷, we calculated for each grid

cell the ratio of the monthly output of the 2×CO₂ simulation to that of the 1×CO₂ simulation for each of the three output variables; temperature was converted to Kelvin before calculating monthly temperature ratios. We then multiplied each ratio by the corresponding variable in our data for contemporary climate to determine the input data for TEM that represent the 2×CO₂ climate for each GCM.

To help separate the effects of changes in CO₂ concentration from those of the GCM climates on estimates of NPP, we did a factorial experiment with TEM involving two levels of CO₂ (312.5 p.p.m.v. and 625.0 p.p.m.v.) and five climate scenarios (contemporary and the four GCM climates). We chose the CO₂ level of 312.5 p.p.m.v. because it was the average baseline concentration of the four GCMs (range: 300–326 p.p.m.v.).

NPP responses to doubled CO₂

For doubled CO₂ with no climate change, TEM predicts a global NPP increase of 16.3% (Table 3). The responses differ widely

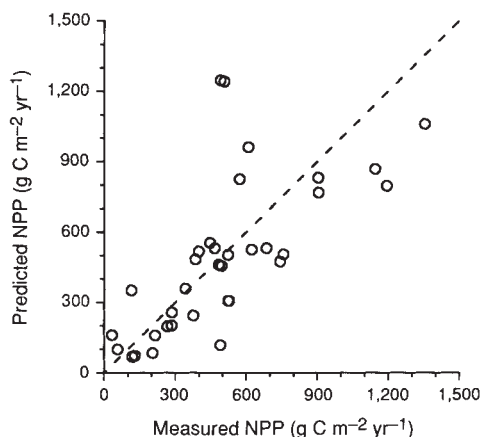


FIG. 3 Comparison between the annual NPP predictions of the TEM and the NPP field measurements that were used to construct several regression-based global models of NPP⁵ ($r=0.69$, $N=35$, $P<0.001$). The dashed line indicates equality between predicted and measured NPP.

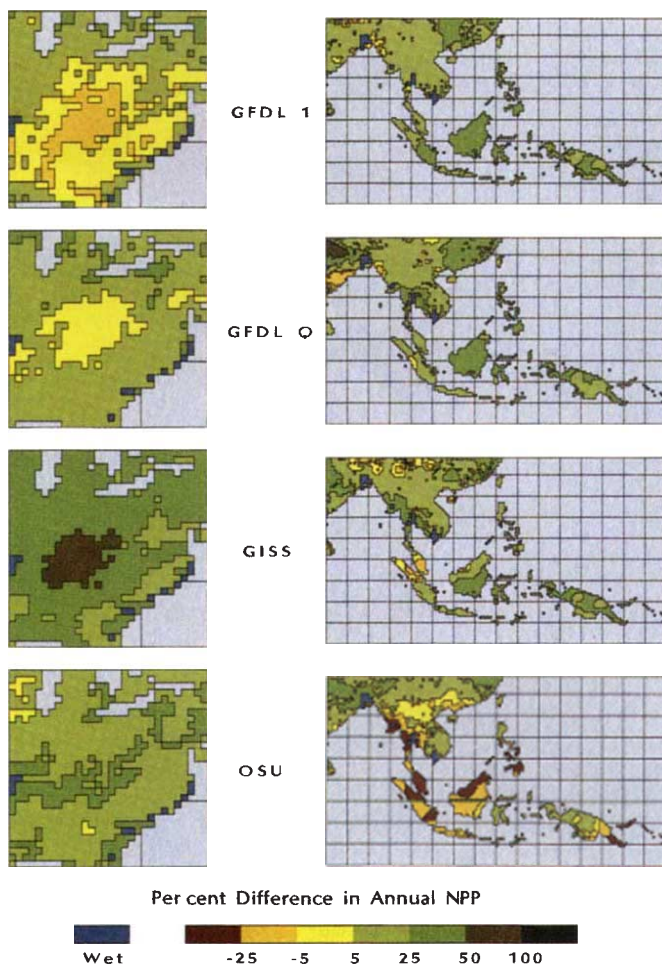


FIG. 4 Per cent difference in annual NPP between contemporary climate at 312.5 p.p.m.v. CO₂ and the various GCM climates at 625.0 p.p.m.v. CO₂ as predicted by the TEM for eastern North America (left) and southeast Asia (right). Estimates were not made for open water (light blue) or wetland ecosystems (bright blue).

among vegetation types and range from no increases for some northern ecosystems to increases of 50.0% for deserts. In general, responses for northern and temperate ecosystems are less than 10%. Responses for many dry ecosystems, such as deserts, arid shrublands, and xeromorphic forests, are over 20%. Tropical evergreen forests, which experience an increase of 22.2% or 4.0 Pg C per year, account for about half of the global increase.

In many northern and temperate ecosystems, NPP is known to be limited by the availability of inorganic nitrogen in the soil⁴⁸. Because of nitrogen limitation, TEM predicts that these ecosystems have low capacity to incorporate elevated CO₂ into production⁷. In nutrient-limited systems, the response of plant growth to elevated CO₂ is often constrained under conditions of low nutrient availability^{49–56}.

Under contemporary climate conditions in dry regions, TEM generally predicts that water availability limits productivity more than nitrogen availability⁸. With a doubling of atmospheric CO₂, TEM predicts that the water-use efficiency of vegetation increases⁸ in these regions; that is, there will be an increase in production per amount of water used. These predictions are consistent with research findings that increases in water-use efficiency with elevated CO₂ are generally greatest in water-stressed systems⁵⁷.

The TEM generally estimates that nitrogen limitation of NPP is much weaker in tropical forests than in temperate and boreal forests¹². Thus, the model predicts that tropical ecosystems are able to incorporate a substantial proportion of elevated CO₂ into production. But, this result must be treated with caution because the response of tropical NPP to elevated CO₂ may be sensitive to some processes that are not presently represented in TEM; these include phosphorus limitation of NPP and C4 photosynthesis. Phosphorus-deficient soils predominate in some regions of the tropics⁵⁸ and most grasses in tropical savannas use the C4 photosynthetic pathway⁵⁹, which is generally less responsive to elevated CO₂ than C3 metabolism⁵⁷. A substantial reduction in NPP response for tropical ecosystems would result in a much lower estimate of global NPP response to doubled CO₂.

NPP responses to changes in climate

Changes in climate with no change in CO₂ concentration are predicted to have little effect on global estimates of NPP (Table 3); the changes in NPP range from a decrease of 2.4% for the OSU climate to essentially no change for the other GCM climates (Table 3). Climate changes may affect NPP in a variety of ways. Elevated temperature may decrease NPP by decreasing soil moisture or enhancing plant respiration. It may also increase NPP by metabolically enhancing photosynthesis or increasing nutrient availability through higher rates of decomposition. In dry regions, lower precipitation or cloudiness may decrease NPP by lowering soil moisture. In moist regions, increased cloudiness may decrease NPP by reducing the availability of PAR. The relative importance of different climate variables in affecting NPP response varies among ecosystems.

For northern and temperate ecosystems, increases in NPP are generally predicted in response to climate change. For all the GCM climates, predicted temperature increases are greatest towards the poles and least in the tropics². Productivity in northern and temperate ecosystems is substantially limited by nitrogen availability, and increases in productivity predicted by TEM are primarily driven by the effect of elevated temperature in enhancing the mineralization of nitrogen in the soils of these regions^{8,12}.

The predicted NPP decreases for tropical evergreen forest, which range between 8.9% and 20.6% among the GCM climates (Table 3), may be related to increased temperature and cloudiness. Increased temperature in TEM may enhance plant respiration enough to decrease NPP in regions where nitrogen availability does not substantially limit production⁸, which is generally predicted by TEM for tropical forests¹². Increases in cloudiness

TABLE 3 Comparison of annual NPP (10^{15} g C) by vegetation type for experiment involving two levels of atmospheric CO_2 and five levels of climate

CO ₂ scenarios Climate scenarios:	312.5 p.p.m.v.					625.0 p.p.m.v.				
	Contemporary	GFDL 1	GFDL Q	GISS	OSU	Contemporary	GFDL 1	GFDL Q	GISS	OSU
Polar desert/alpine										
tundra	0.4	0.4	0.4	0.4	0.4	0.5	0.5	0.5	0.5	0.5
Wet/moist tundra	0.6	0.7	0.7	0.7	0.7	0.6	0.8	0.7	0.7	0.7
Boreal woodland	1.1	1.4	1.3	1.3	1.3	1.1	1.6	1.4	1.4	1.4
Boreal forest	2.9	3.8	3.6	3.6	3.5	2.9	4.4	4.0	3.7	3.7
Temperate coniferous										
forest	1.1	1.1	1.1	1.1	1.1	1.2	1.3	1.3	1.4	1.3
Desert	0.6	0.6	0.5	0.6	0.6	0.9	1.0	0.9	1.0	1.0
Arid shrubland	1.8	1.8	1.8	1.9	1.9	2.3	2.5	2.5	2.7	2.6
Short grassland	1.0	1.2	1.2	1.2	1.1	1.1	1.4	1.3	1.4	1.2
Tall grassland	1.2	1.4	1.4	1.5	1.3	1.3	1.5	1.5	1.6	1.4
Temperate savanna	2.2	2.3	2.4	2.6	2.4	2.5	2.9	2.9	3.1	2.9
Temperate deciduous										
forest	2.2	2.0	2.1	2.5	2.3	2.3	2.4	2.6	2.8	2.6
Temperate mixed forest	3.3	3.4	3.5	3.7	3.6	3.6	4.0	4.1	4.2	4.0
Temperate broadleaf										
evergreen forest	2.2	2.3	2.2	2.2	2.2	2.6	2.8	2.8	2.8	2.7
Mediterranean shrubland	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.7	0.6
Tropical savanna	5.3	5.7	5.6	6.0	6.0	5.6	6.3	6.3	6.7	6.6
Xeromorphic forest	2.9	2.7	2.7	2.7	2.9	3.7	3.7	3.8	3.8	4.1
Tropical deciduous forest	3.8	3.4	3.5	3.3	3.5	4.5	4.5	4.6	4.5	4.6
Tropical evergreen forest	18.0	16.4	16.3	15.6	14.3	22.0	21.9	21.8	21.3	19.3
Total	51.0	51.1	50.8	51.5	49.8	59.3	64.3	63.8	64.2	61.2

Ecosystem-based estimates may not sum to totals because of the effects of rounding in reporting those estimates.

in tropical evergreen forest may decrease PAR enough to decrease NPP¹⁸. The largest NPP decrease for tropical evergreen forest occurs for the OSU climate which predicts the largest increase in mean annual cloudiness (9.8%).

NPP responses to changes in CO₂ and climate

Compared to global NPP for contemporary climate at 312.5 p.p.m.v. CO₂, the global responses to changes in both CO₂ and climate do not vary substantially among the GCM climates with increases ranging between 20.0% and 26.1% (Table 3). For tropical and dry temperate ecosystems, increases in NPP are dominated by the effects of elevated CO₂. But for northern and moist temperate ecosystems, NPP increases reflect primarily the effects of elevated temperature in enhancing nitrogen availability.

Although the predicted global and ecosystem-wide responses of NPP at elevated CO₂ are generally similar for all the GCM climates, there are differences in NPP response at smaller spatial scales. For example, in eastern North America the NPP decreases just west of the Appalachian Mountains for the GFDL 1 climate (Fig. 4) may be caused by decreased precipitation. The GFDL 1 climate predicts a decrease of 7.0% in annual precipitation for temperate mixed forest in the region. In southeast Asia, the substantial NPP decreases in Indonesia for the OSU climate (Fig. 4) may be caused by increased cloudiness which reduces PAR. The OSU climate predicts an increase of 12.5% in mean annual cloudiness for tropical evergreen forest in the region.

Conclusion

The application of TEM in this study demonstrates our ability to explore, in a mechanistic manner, the potential consequences of changes in CO₂ and climate for NPP across the entire terrestrial surface of the globe. Our results indicate that simultaneous interactions among the dynamics of carbon, nitrogen, and water affect the ability of vegetation to incorporate elevated CO₂ into production. Because these interactions are both complex and spatially variable, assessments of how changes in CO₂ and climate affect NPP require the use of process-based models that are geographically referenced. The TEM represents an important

tool for making these assessments because it provides scientists and policy makers with the capability to investigate the potential effects of climate change on biospheric functions in a quantitative and geographically specific way.

Although the results of this study represent our current understanding of how changes in CO₂ and climate will affect terrestrial NPP, they do not consider the redistribution of vegetation that may result from climate change. Future studies with TEM will evaluate the sensitivity of NPP to vegetation redistribution in addition to changes in CO₂ and climate. □

Received 9 November 1992; accepted 29 March 1993.

- Watson, R. T., Filho, L. G. M., Sanhueza, E. & Janetos, A. in *Climate Change 1992: The Supplementary Report to the IPCC Scientific Assessment* (eds Houghton, J. T. et al.) 25–46 (Cambridge Univ. Press, Cambridge, 1992).
- Mitchell, J. F. B., Manabe, S., Meleshko, V. & Tokioka, T. in *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T. et al.) 131–172 (Cambridge Univ. Press, Cambridge, 1990).
- Melillo, J. M., Callaghan, T. V., Woodward, F. I., Salati, E. & Sinha, S. K. in *Climate Change: The IPCC Scientific Assessment* (eds Houghton, J. T. et al.) 283–310 (Cambridge Univ. Press, Cambridge, 1990).
- Agren, G. I., McMurtrie, R. E., Parton, W. J., Pastor, J. & Shugart, H. H. *Ecol. Applic.* **1**, 118–138 (1991).
- Lieth, H. in *Primary Productivity of the Biosphere* (eds Lieth, H. & Whittaker, R. H.) 237–263 (Springer, New York, 1975).
- Esser, G. *Tellus* **39B**, 245–260 (1987).
- Esser, G. in *Soils and the Greenhouse Effect* (ed. Bouwman, A. F.) 249–261 (Wiley, Chichester, 1990).
- McGuire, A. D. et al. *Clim. Change* (in press).
- Schimel, D. S., Parton, W. J., Kittel, T. G. F., Ojima, D. S. & Cole, C. V. *Clim. Change* **17**, 13–25 (1990).
- Burke, I. C. et al. *BioScience* **41**, 685–692 (1991).
- Running, S. W. & Nemani, R. R. *Clim. Change* **19**, 349–368 (1991).
- McGuire, A. D. et al. *Glob. biogeochem. Cycles* **6**, 101–124 (1992).
- Emanuel, W. R., Shugart, H. H. & Stevenson, M. P. *Clim. Change* **7**, 29–43 (1985).
- Prentice, K. C. *J. geophys. Res.* **95** (D8), 11811–11830 (1990).
- Woodward, F. I. & McKee, I. F. *Envir. Int.* **17**, 535–546 (1991).
- Prentice, I. C. et al. *J. Biogeogr.* **19**, 117–134 (1992).
- Smith, T. M., Leemans, R. & Shugart, H. H. *Clim. Change* **21**, 367–384 (1992).
- Raich, J. W. et al. *Ecol. Applic.* **1**, 399–429 (1991).
- Kimball, B. A. *Agronomy J.* **75**, 779–788 (1975).
- Gates, D. M. in *Direct Effects of Increasing Carbon Dioxide on Vegetation, Report DOE/ER-0238* (eds Strain, B. R. & Cure, J. D.) 171–184 (US Department of Energy, Washington DC, 1985).
- White, F. *Vegetation of Africa* (UNESCO, Paris, 1981).
- Institute of Geography of the Siberian Department of USSR Academy of Sciences, Botanical Institute of USSR Academy of Sciences & Moscow State University Geography Department *Vegetation of the USSR* (GVGK, Minsk, 1990).
- Hou, H. Y. et al. *Vegetation Map of China* (Map Publisher of the People's Republic of China, Beijing, 1979).

24. Matthews, E. *J. Clim. Appl. Meteorol.* **22**, 474–487 (1983).
25. Olson, J. S., Watts, J. A. & Allison, L. *J. Carbon in Live Vegetation of Major World Ecosystems*, Environmental Sciences Division Publication No. 1997 (Oak Ridge National Laboratory, Oak Ridge, Tennessee, 1983).
26. Moscow State University Geographical Department *Geographical Belts and Zonal Types of the Landscapes of the World Map* (GVGK, Moscow, 1988).
27. Australian Surveying and Land Information Group *Vegetation. Atlas of Australian Resources, Third Series*, Vol. 6 (Commonwealth of Australia, Canberra, 1990).
28. Rowe, J. S. *Forest Regions of Canada Publ. 1300* (Department of Fisheries and the Environment, Can. Forest Serv., Ottawa, 1972).
29. Joint Federal-State Land Use Planning Commission for Alaska *Major Ecosystems of Alaska* (US Geological Survey, Fairbanks, Alaska, 1973).
30. Kuchler, A. W. *Potential Natural Vegetation of the Conterminous United States* (American Geographical Society, New York, New York, 1964).
31. UNESCO *Vegetation Map of South America* (UNESCO, Paris, 1981).
32. Vorosmarty, C. J. *et al. Globl biogeochem. Cycles* **3**, 241–265 (1989).
33. Deevey, E. S. Jr *Scient. Am.* **203**, 195–204 (1960).
34. Whittaker, R. H. *Communities and Ecosystems* (Macmillan, New York, 1970).
35. Olson, J. S. in *Temperate Forest Ecosystems* (ed. Reichle, D. E.) 226–241 (Springer, New York, 1970).
36. Bazilevich, N. I., Rodin, L. E. & Rozov, N. N. *Untersuchungen der Biologischen Produktivität in Geographischer Sicht* (5th Tagung Geogr. Ges., Leningrad, 1970).
37. Study of Critical Environmental Problems (SCEP) *Man's Impact on the Global Environment* (MIT Press, Cambridge, Massachusetts, 1970).
38. Golley, F. B. in *Ecosystem Structure and Function*, Ann. Biol. Colloq. **31** (ed. Wiens, J. A.) 69–70 (Oregon State University, Corvallis, Oregon, 1972).
39. Whittaker, R. H. & Likens, G. E. *Hum. Ecol.* **1**, 357–369 (1973).
40. Whittaker, R. H. & Likens, G. E. in *Primary Productivity of the Biosphere* (eds Lieth, H. & Whittaker, R. H.) 305–328 (Springer, New York, 1975).
41. Ajtay, G. L., Ketner, P. & Duvinéaud, P. in *The Global Carbon Cycle SCOPE 13* (eds Bolin, B., Degens, E. T., Kempe, S. & Ketner, P.) 129–182 (Wiley, Chichester, 1979).
42. Heimann, M. & Keeling, C. D. in *Aspects of Climate Variability in the Pacific and Western Americas*, Geophysical Monograph **55** (ed. Peterson, D. H.) 237–275 (Amer. Geophys. Union, Washington DC, 1989).
43. Schlesinger, W. H. *Biogeochemistry: An Analysis of Global Change* (Academic, San Diego, California, 1991).
44. Paul, E. A. & Clark, F. E. *Soil Microbiology and Biochemistry* (Academic, San Diego, California, 1989).
45. Jenne, R. L. in *Global Climate Change: Implications, Challenges and Mitigation Measures* (eds Majumdar *et al.*) 145–164 (Pennsylvania Academy of Sciences, Easton, Pennsylvania, 1992).
46. Willmot, C. J., Rowe, C. M. & Philpot, W. D. *J. Am. Cart.* **12**, 5–16 (1985).
47. Adams, R. M. *et al. Nature* **345**, 219–224 (1990).
48. Vitousek, P. M. & Howarth, R. W. 1991. *Biogeochemistry* **13**, 87–115 (1991).
49. Wong, S. C. *Oecologia* **44**, 68–74 (1979).
50. Larigauderie, A., Hilbert, D. W. & Oechel, W. C. *Oecologia* **77**, 544–549 (1988).
51. Goudriaan, J. & de Ruiter, H. E. *Neth. J. agric. Sci.* **31**, 157–169 (1983).
52. Zangerl, A. R. & Bazzaz, F. A. *Oecologia* **62**, 412–417 (1984).
53. Brown, K. & Higginbotham, K. O. *Tree Physiol.* **2**, 223–232 (1986).
54. Oberbauer, S. F., Sionit, N., Hastings, S. J. & Oechel, W. C. *Can. J. Bot.* **64**, 2993–2998 (1986).
55. Marks, S. & Clay, K. *Oecologia* **84**, 207–214 (1990).
56. Johnson, R. H. & Lincoln, D. E. *Oecologia* **87**, 127–134 (1991).
57. Mooney, H. A., Drake, B. G., Luxmoore, R. J., Oechel, W. C. & Pitelka, L. F. *BioScience* **41**, 96–104 (1991).
58. Sanchez, P. A., Bandy, E. E., Villachica, J. H. & Nicholaides, J. J. *Science* **216**, 821–827 (1982).
59. Long, S. P. & Hutchin, P. R. *Ecol. Appl.* **1**, 139–156 (1991).

ACKNOWLEDGEMENTS. This work was primarily funded by the National Aeronautics and Space Administration. Other support came from the USDA Forest Service Resources Program and Assessment Staff, which funded A. D. McG. The National Science Foundation and the Andrew W. Mellon Foundation. D. Martin and L. Thomson of the Marine Biological Laboratory produced the colour plates. S. Mazurkiewicz of the Complex Systems Research Center assisted with processing the climate data. We thank J. D. Aber, C. Field, E. A. Griffin, A. Janetos, I. C. Prentice, J. W. Raich, C. Rosenzweig and F. I. Woodward for comments on the manuscript.

A measurement of the optical depth through a galaxy disk

P. A. James & P. J. Puxley

Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK

SEVERAL recent papers^{1–3} have refocused attention on the question of the degree of dust obscuration (or optical depth) of the disks of spiral galaxies. The traditional view, based on the early statistical studies of Holmberg⁴ and de Vaucouleurs⁵, is that the disks are largely transparent, whereas Disney *et al.*¹ and Valentijn² have recently argued in favour of substantial optical depths. In an attempt to resolve this issue, White and Keel³ measured the broad-band colours of an overlapping galaxy pair, to determine directly the optical depth of the foreground galaxy; however, using this method it was not possible to separate unambiguously the contributions from the two galaxies. Individual emission lines, on the other hand, are completely separated if the redshift difference between the two galaxies is large enough, as in the pair studied here, offering a means of removing the source ambiguity. By comparing the $H\alpha/H\beta$ ratios for the H II regions in the background galaxy with those in isolated spiral galaxies, and exploiting the wavelength dependence of extinction by dust, we find significant optical depths along three lines of sight through the foreground galaxy.

Our analysis is based on observations of NGC3314a and b, two spiral galaxies along the same line of sight, which are described fully by Schweizer and Thonnard⁶. An excellent image of the system is shown in ref. 7. The nearer of the two, NGC3314a, is a face-on Sc, with a redshift $z=0.0095$ ($2,835 \text{ km s}^{-1}$), through the disk of which can be seen NGC3314b, a probable Sb with an inclination of $\sim 65^\circ$ and $z=0.0155$ ($4,641 \text{ km s}^{-1}$). Spectroscopy shows that the inclined galaxy, with a correspondingly large rotation amplitude, is the more highly redshifted, and thus is likely to be the background object, barring large non-Hubble velocities. Both galaxies appear undisturbed, indicating that they are well separated in space. They are remarkably well aligned: in ref. 8 the nucleus of NGC3314b is resolved 2–3 arcsec northwest of that of NGC3314a.

We used the Balmer decrement, $H\alpha/H\beta$, of emission from H II regions (ionized gas surrounding hot stars) in the background galaxy, to determine the optical depth through NGC3314a. The redshift difference of the two galaxies com-

pletely separates their emission in these lines. This is the main advantage of this spectroscopic approach over that of Keel⁸ and White and Keel³, who used broad-band colours to tackle the same problem, for NGC3314 and another overlapping galaxy pair, but who had no way to separate unambiguously the emission from the foreground and background galaxies. Furthermore, the nature of this galaxy pair means that we are sampling extinction right through a galaxy disk, removing the uncertainties about relative scale heights of dust and stars which affect studies of surface brightness as a function of inclination.

The principal weakness of the method is that it can only be applied to this one object. We know of no other suitable galaxy pairs, and the observations presented here identify only four lines of sight through NGC3314a backlit by Balmer line emission from NGC3314b. A further worry is a possible selection bias towards low extinction regions, as no emission would be detected through areas of very heavy obscuration in the foreground galaxy. The visual appearance of the galaxy provides some reassurance that this problem is not great: the background galaxy can be seen clearly through most of the disk of NGC3314a, indicating that the observed emission is not coming through a few holes in an otherwise optically thick disk.

The spectra presented here were obtained at the University of Hawaii 2.24-m telescope on Mauna Kea, on 1991 January 10 and 11. We used the facility Faint Object Spectrograph with the NSF1 800 × 800 CCD (charge-coupled device), using a 1,200 line mm^{-1} grating for the $H\alpha$ observations, and a 600 line mm^{-1} grating at $H\beta$. A 275 μm -wide slit was used, corresponding to 6.25-Å resolution at 6,700 Å, sufficient to resolve $H\alpha$ and N II, and subtending 2.5 arcsec on the sky. The slit was aligned with the major axis of NGC3314b, 140° east of north, and was centred on the optical nucleus. There was no attempt to position the slit on specific H II regions in the background galaxy. Integration times were 105 minutes at $H\alpha$ and 240 minutes at $H\beta$, and the conditions were photometric throughout. A Cu–Ar arc was used for wavelength calibration, and two stars, Feige 34 and Feige 67, from the list of Massey *et al.*⁹, were observed for spectro-photometric calibration.

$H\alpha$ emission was detected from the nucleus and three H II regions in NGC3314b. Two H II regions were northwest of the nucleus, by 10.5 and 26 arcsec, and one 11.5 arcsec to the southeast. These correspond to radial distances of 1.45, 3.6 and 1.6 $h^{-1} \text{ kpc}$ at the distance of NGC3314a, where h is the Hubble constant in units of $100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and assuming that the measured redshift of NGC3314a is purely cosmological. $H\beta$

TABLE 1 Line fluxes and extinction coefficients

Region	H α	H β	$A_{V,tot}$	$A_{V,ext}$
1	55.6	10.0	1.72 ± 0.33	1.58 ± 0.33
2	105.7	17.4	1.94 ± 0.33	1.80 ± 0.33
3	62.8	<7.9	>2.64	>2.5
4	31.7	6.0	1.58 ± 0.34	1.44 ± 0.34

Fluxes are in units of $10^{-19} \text{ W m}^{-2}$. $A_{V,tot}$ is total extinction; $A_{V,ext}$ is extinction external to the Galaxy.

emission was clearly detected from all three off-nuclear regions, but only marginally from the nucleus.

The CCD spectroscopy was reduced and calibrated using standard techniques. The H β line was found to lie in a deep absorption feature, arising from photospheric H β in the stars of NGC3314b. This was removed by fitting and subtracting the continuum from a neighbouring region of the galaxy, which had no H β emission. No absorption was apparent at H α , and it must be significantly weaker than at H β , possibly because of diffuse H α line emission from disk gas filling in the absorption line. No correction was applied, but if the H α absorption equivalent width were equal to that at H β , the emission line strengths would be increased by 10–15%. This would have the effect of increasing the extinction values derived below, but not by more than the quoted errors.

Line fluxes were measured using our own software package, which fitted the spectra by a gaussian emission line plus a sloping continuum. We derived the best-fit linewidths (10.03 Å full width at half maximum (FWHM) at H α , 4.54 Å FWHM at H β) from the H II region with the best signal-to-noise ratio, and fixed the widths at these values for the fits to the remaining regions. The line fluxes are listed in Table 1, where the numbering runs northwest-southeast, and the nucleus is thus region 3. The errors were assumed to be dominated by calibration uncertainties, for which we adopted 10% at both H α and H β . The H β flux from the nucleus of NGC3314b is a marginal detection, with the line only becoming apparent after application of the absorption correction described above. Thus we treat the flux obtained as an upper limit.

The line ratios were converted to extinction values by assuming that the intrinsic ratios are optically thick to all Lyman

transitions (Case B), with an electron density of 100 cm^{-3} and temperature of 10^4 K (ref. 10), giving an emitted ratio of 2.86. Assuming a standard extinction law¹¹, which includes the effects of absorption and scattering within the dust screen, this gives the following relation between extinction and the observed line ratio

$$A_V = 2.6 \times \ln \frac{(F_{H\alpha}/F_{H\beta})_{OBS}}{2.86} \quad (1)$$

This equation was used to calculate the total extinction between the telescope and the H II regions in NGC3314b, listed in column 4 of Table 1. But there are three contributions to these values: from dust within NGC3314b, through the disk of NGC3314a, and within the disk of the Galaxy. We are only interested in the second of these. The galactic component¹² corresponded to A_V of 0.14. Subtracting this value gives the extinction external to the Galaxy, which is listed in the final column of Table 1. We are only able to put a lower limit on the nuclear extinction, of $A_V > 2.5$ magnitudes, because of the non-detection of H β at a reliable level. The remaining analysis will concern only the extinction values obtained for the three off-nuclear sources.

Accounting for the extinction internal to the source galaxy is complicated, as there will be a range of extinction values to the different H II regions. We thus did a statistical test, comparing the extinction values towards NGC3314b with those for H II regions in well studied galaxies from the literature. This method also accounts for the effects of scattering within the source galaxy. Belley and Roy¹³ present CCD narrow-band photometry at H α and H β for 130 H II regions in NGC628 and 160 H II regions in NGC6946. We calculated A_V values for each of the H II regions, using the same method as for NGC3314, again subtracting a galactic extinction component¹². For NGC6946, this correction is large, and as a result about 40 of the H II regions had negative corrected A_V values. These were ignored in the statistical comparison with NGC3314, a conservative procedure as, for example, including them at their calculated value, or setting them to an A_V of zero, would have increased the significance of the difference presented below.

The corrected A_V distributions for NGC628 and NGC6946 both have the bulk of their points between 0 and 0.5 magnitudes, with a high- A_V tail stretching to 1.5 magnitudes. However, all three of the off-nuclear H II regions in NGC3314b have A_V values of > 1.4 , and thus seem unlikely to come from the same distribution (Fig. 1). This was quantified by doing a one-sample Kolmogorov-Smirnov test, using the data for NGC628 and NGC6946 to define parent distributions, and calculating the probability that the values for NGC3314b were drawn from these distributions. These probabilities were found to be 2×10^{-4} and 8×10^{-4} respectively, with the larger value for NGC6946 probably reflecting the conservatively ignored negative A_V values. In both cases, however, it is clear that the extinction values towards NGC3314b are significantly higher than for the comparison galaxies.

The low A_V values towards the H II regions in NGC628 and NGC6946 are presumably due to a low dust scale height in these galaxies, enabling giant H II regions to 'burst through', and be seen relatively unaffected by extinction (J.-R. Roy, personal communication). These values are thus not representative of the extinction right through the disks of these galaxies.

NGC3314b is highly inclined, which could contribute to the extinction. However, Garnett and Shields¹⁴ give line fluxes for 18 H II regions in M81, which has an inclination of 60° , and yields A_V values entirely consistent with the distributions for NGC628 and NGC6946. Although a sample of 18 is too small for a meaningful statistical comparison with the three regions of NGC3314b, the largest A_V of the 18 was only 1.04, much smaller than the values for NGC3314b. Roughly half of the H II regions in M81 are at radial distances from the nucleus that are similar to the three in NGC3314b, and should thus be directly comparable. There is no significant trend in extinction with

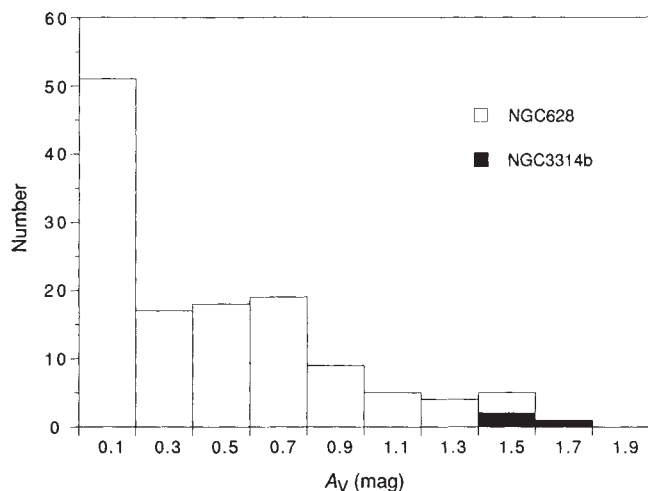


FIG. 1 The distribution of visual extinction (A_V) for 130 H II regions in NGC628 (unshaded), from Belley and Roy¹³, compared with the three H II regions in NGC3314b (solid shading).

radius for M81, again as is found for NGC628 and NGC6946. NGC3314b is slightly more inclined than M81, and thus we may still be underestimating the internal extinction, and overestimating that through NGC3314a.

The excess visual extinction towards NGC3314b seems to be at least 0.8 magnitudes for marginal statistical consistency, with the best-fit value being much larger, 1.3 magnitudes. The obvious interpretation is that this excess represents the extinction through the disk of NGC3314a. This value, corresponding to an A_B of about 1.7 magnitudes, is in general agreement with the large extinction values preferred by Disney *et al.*¹ and Valentijn². We are looking through regions of the foreground galaxy that are well out into the disk, as $2\text{--}4\text{ h}^{-1}\text{ kpc}$ corresponds to $1\text{--}2$ disk scale lengths for a typical spiral. Valentijn's results require significant optical depths out to greater radii than we are probing here, so our measured extinctions, although large, are not in unequivocal agreement. Of the uncertainties in our result, only inclination effects act so as to decrease the inferred extinction, whereas both possible $H\alpha$ absorption and selection towards 'holes' in NGC3314a act in the opposite sense. Thus we feel that our estimates are conservative. $\square$

Received 11 January; accepted 30 March 1993.

1. Disney, M., Davies, J. & Philipps, S. *Mon. Not. R. astr. Soc.* **239**, 939–976 (1989).
2. Valentijn, E. A. *Nature* **346**, 153–155 (1990).
3. White, R. E. & Keel, W. C. *Nature* **359**, 129–131 (1992).
4. Holmberg, E. *Medd. Lund. Ser. II*, No. 136 (1958).
5. de Vaucouleurs, G. *Astr. J.* **64**, 397–409 (1959).
6. Schweizer, F. & Thonnard, N. *Publ. astr. Soc. Pacific* **97**, 104–109 (1985).
7. Richter, O.-G., Materne, J. & Huchtmeier, W. K. *Astr. Astrophys.* **111**, 193–200 (1982).
8. Keel, W. C. *Astr. J.* **88**, 1579–1586 (1983).
9. Massey, P., Strobel, K., Barnes, J. V. & Anderson, E. *Astrophys. J.* **328**, 315–333 (1988).
10. Hummer, D. G. & Storey, P. J. *Mon. Not. R. astr. Soc.* **224**, 801–820 (1987).
11. Savage, B. D. & Mathis, J. S. A. *Rev. Astr. Astrophys.* **17**, 73–111 (1979).
12. Burstein, D. & Heiles, C. *Astrophys. J. Suppl.* **54**, 33–79 (1984).
13. Belley, J. & Roy, J.-R. *Astrophys. J. Suppl.* **78**, 61–85 (1992).
14. Garnett, D. R. & Shields, G. A. *Astrophys. J.* **317**, 82–101 (1987).

Magnetic resonance of a single molecular spin

J. Köhler, J. A. J. M. Disselhorst, M. C. J. M. Donckers, E. J. J. Groenen, J. Schmidt† & W. E. Moerner*†

Centre for the Study of Excited States of Molecules, Huygens Laboratory, University of Leiden, PO Box 9504, 2300 RA Leiden, The Netherlands

* IBM Research Division, Almaden Research Center, 650 Harry Road, San Jose, California 95120-6099, USA

THE introduction of optical detection methods for observing magnetic resonance transitions in metastable paramagnetic states^{1–4} has contributed enormously to our understanding of the properties of photoexcited molecules in condensed phases. In such experiments the luminescence intensity is recorded as a function of the frequency of an applied microwave field. At resonance with transitions between sublevels of a metastable paramagnetic state, the lifetime of the metastable state is altered and a consequent change in the luminescence intensity is observed. Here we report the observation of such optically detected magnetic resonance transitions for the triplet state of a single pentacene molecule embedded in a *p*-terphenyl host crystal. This result has been obtained by combining the conventional optical detection technique for observing magnetic resonance transitions^{1–4} with the new single-molecule optical detection methods developed recently^{5,6}. This observation opens the way for magnetic resonance studies in condensed phases with single-molecule sensitivity.

Sublimation-grown single crystals of *p*-terphenyl containing

$10^{-7}\text{--}10^{-8}$ mol/mol pentacene were cooled to 1.5 K. As described in detail previously⁷, fluorescence excitation of single pentacene molecules in these samples was achieved by focusing a narrow-band single-mode laser to a spot of diameter $5\text{ }\mu\text{m}$ and by tuning the laser far into the wing of the inhomogeneously broadened $S_1 \leftarrow S_0$ electronic singlet-singlet transition (this broadening is due to the slightly different environments of the guest molecules) such that only one pentacene molecule was in resonance. The emitted fluorescence shifted to longer wavelengths was collected by a parabolic mirror and recorded with a GaAs photomultiplier and a photon counter. To obtain the magnetic-resonance signals, the laser was tuned exactly to the top of the homogeneously broadened absorption line of a single pentacene molecule and the intensity adjusted to provide an optimum average triplet-state population. A microwave field (amplitude modulated at 37 Hz, typical power 100 mW) was applied with a two-turn, grounded helix around the sample. The photon counter was operated in a two-channel gated mode, one channel accumulating counts while the microwaves were off and the other channel while they were on. The fluorescence-detected magnetic-resonance (FDMR) signal, which corresponds to the difference between these two channels, was recorded with a count interval of 10 s as a function of the microwave frequency.

Figure 1 shows the FDMR spectrum of the triplet state of a single molecule of pentacene in a crystal of *p*-terphenyl. The detected fluorescence (inset of Fig. 1) corresponds to the radiative decay from S_1 on continuously pumping the $S_1 \leftarrow S_0$ transition of the single molecule (see Fig. 2). If the molecule escapes from this cycle into the lowest triplet state T_1 , it most probably ends up in the T_x sublevel because of the selectivity of the intersystem crossing process⁸. On saturating the $T_x \rightarrow T_z$ transition with microwaves, the mean residence time of the molecule in the triplet state lengthens by about a factor of 2, given the lifetimes of the triplet sublevels⁸. This shows up as a decrease in the number of fluorescence counts per time interval, the FDMR signal. When the laser is fixed in resonance with the single molecule at frequency ν_1 (Fig. 1 inset), the FDMR signal

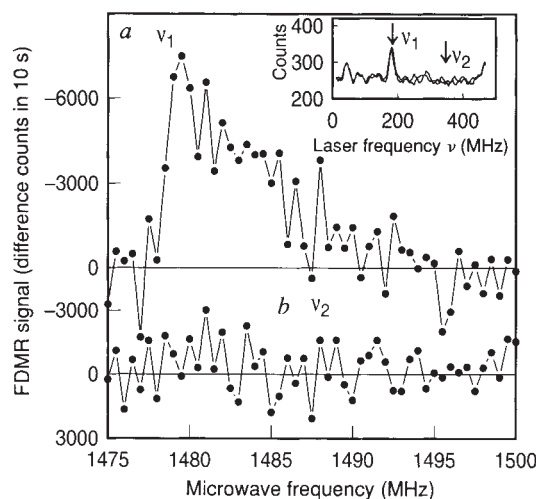


FIG. 1 a, FDMR signal of a single pentacene molecule in a *p*-terphenyl host crystal. The molecule was excited at the fixed laser frequency ν_1 (equivalent to 592.404 nm). This is indicated in the inset, where two consecutive fluorescence excitation spectra obtained by scanning the laser are shown. b, No FDMR signal is visible when the laser frequency is held at ν_2 . The dots give the measured values, and the lines serve as a guide to the eye. The average photon-count rate during acquisition of trace a was $\sim 2 \times 10^4\text{ s}^{-1}$, and the laser intensity was $\sim 200\text{ mW cm}^{-2}$. The spectra in the inset were recorded with a lower laser intensity than that used for the FDMR signals.

† To whom correspondence should be addressed.

of Fig. 1a is observed. If the frequency is fixed at ν_2 , out of resonance with any molecule, no FDMR signal is observed (Fig. 1b). This is clear evidence that we have observed the magnetic-resonance transition of a single molecular spin.

The T_x - T_z transitions observed for pentacene in *p*-terphenyl in a conventional '*N*-molecule' FDMR experiment, where the excitation is provided by a broad-band (1 cm^{-1} bandwidth) dye laser, are shown in Fig. 3a and b. In these experiments, about 10^{13} pentacene molecules are contained in the irradiated part of the crystal. The pentacene molecule occupies four different substitutional sites in the crystal, which produces four distinct origins for the optical transition, named O_1 to O_4 in order of decreasing wavelength. In our samples, these four inhomogeneously broadened lines are well separated. The excitation wavelengths are chosen to select pentacene molecules either in site O_1 (Fig. 3a) or in site O_2 (Fig. 3b). The FDMR signals for these sites show similar, asymmetrical line shapes, which are shifted with respect to each other by about 2 MHz because of a difference in the T_x - T_z splitting produced by the different crystal field strengths in O_1 and O_2 . Single-molecule FDMR signals obtained for a number of different excitation wavelengths are represented in Fig. 3c to g. We have selected excitation wavelengths in the long-wavelength wing of O_1 (molecules I, II, III) and the short-wavelength wing of O_2 (molecule IV). However, comparison of the '*N*-molecule' and the single-molecule spectra suggests that molecule II is in the O_1 site, whereas molecules I, III and IV seem to have a local surrounding corresponding to the O_2 site. The intriguing possibility that, from the point of view of magnetic interactions, O_2 -like molecules may appear in the long-wavelength edge of O_1 must be proven by further work. It is to be expected that the crystal field for molecules in such distorted or strained sites could differ from the crystal fields for molecules at the centre of the inhomogeneously broadened line.

Interestingly, the line shapes in the single and '*N*-molecule' experiments are virtually identical. This observation is in agreement with the ergodic theorem that the time average of an observable for a single molecule is equivalent to the ensemble average of that observable for a system of N molecules. To understand the line shape in the *N*-molecule experiment, we have to consider the hyperfine interaction of the triplet spin

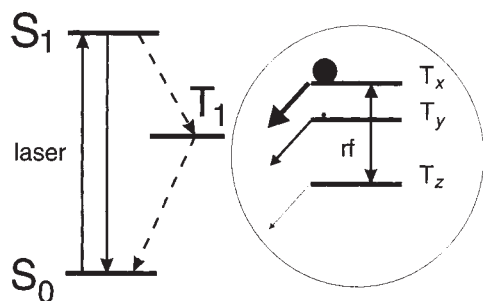


FIG. 2 Energy-level diagram for pentacene. S_0 and S_1 denote the ground state and the first excited singlet state. The arrows indicate the excitation-emission cycle between these states produced by the laser. T_1 is the lowest triplet state which is populated as well as depopulated by radiationless intersystem crossing processes indicated by the broken arrows. The quantum yield of intersystem crossing from S_1 to T_1 is ~ 0.005 for the O_1 and O_2 sites of pentacene in *p*-terphenyl⁹. The rightmost part of the figure shows the triplet state on an enlarged scale. T_x , T_y and T_z are the triplet sublevels split in energy by the magnetic dipole-dipole interaction of the unpaired-electron spins (the zero-field splitting). The arrow between T_x and T_z shows the effect of the applied microwave (rf) radiation. For pentacene in naphthalene, the relative population probabilities of the triplet sublevels are $P_x:P_y:P_z=10:1:0$, as schematically indicated by the circle, and the sublevel lifetimes are $\tau_x=15\text{ }\mu\text{s}$, $\tau_y=35\text{ }\mu\text{s}$ and $\tau_z=280\text{ }\mu\text{s}$ (ref. 8). These values are expected to be fairly similar for pentacene in *p*-terphenyl.

with the 14 protons on the pentacene molecule. The 2^{14} possible nuclear spin states lead to a spread of triplet-spin-resonance frequencies of the order observed in Fig. 3a. In zero magnetic field, the hyperfine interaction between the electron spin of the molecule and the proton spins only gives matrix elements between different electronic substates and thus makes a second-order contribution to the line shape. The question is why, for the single-molecule experiment, the triplet spin does not experience only one of the possible nuclear configurations. Apparently, it 'sees' all of these configurations during the many pumping cycles of the pentacene molecule that occur during the 10 s required for accumulation of sufficient signal-to-noise ratio. When the triplet spin is created, the 14 proton spins on pentacene suddenly feel the (second-order) hyperfine fields, which shift their resonance frequency away from the dipolar spectrum of the protons in the bulk of the crystal. Consequently, during the triplet life time, this surrounding is frozen and the resonance frequency of the triplet spin can only vary in a small interval $\Delta\nu_{\text{loc}}$ determined by the flip-flop motions of the protons in the bulk. This interval can be estimated from the electron spin-spin relaxation time T_2 (ref. 8) and amounts to $\Delta\nu_{\text{loc}}=1/\pi T_2\approx 150\text{ kHz}$. On return to the ground state, the hyperfine fields disappear and the pentacene protons are free to participate in the nuclear flip-flop motion. When the molecule is again excited into T_1 , a new magnetic surrounding is frozen, which corresponds to a different position in the zero-field resonance line. An estimate of the related timescales shows that this process indeed offers a plausible explanation for the observed line shape. In our experiment the average time between two excitations into

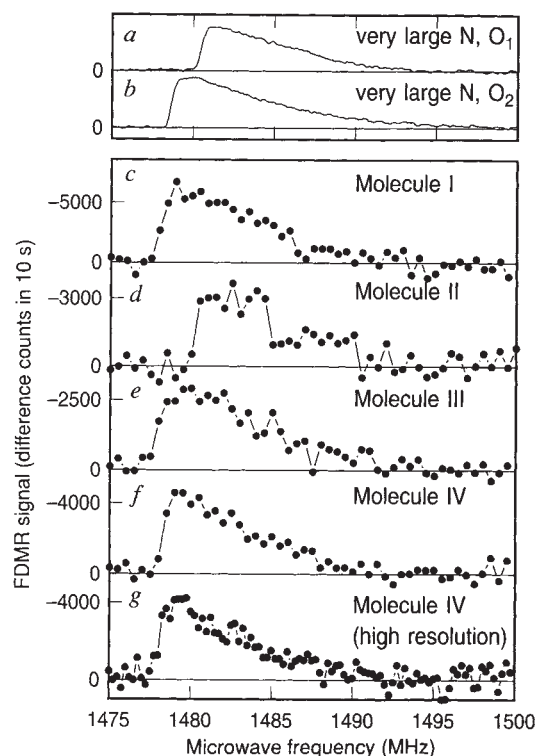


FIG. 3 T_x - T_z transition of pentacene in *p*-terphenyl detected with FDMR for various experimental conditions. a, b, Conventional FDMR spectra for a Bridgman-grown crystal (3×10^{-5} mol pentacene per mol *p*-terphenyl) with excitation in the O_1 and O_2 sites, respectively. c-g, Single-molecule FDMR spectra for a thin sublimed crystal with different excitation wavelengths of the laser ($\lambda=592.447\text{ nm}$ (c), $\lambda=592.370\text{ nm}$ (d), $\lambda=592.404\text{ nm}$ (e), $\lambda=592.065\text{ nm}$ (f, g)). Traces c and e are averages of three, trace f of eight and trace g of four individual scans. Spectra a and b are given in arbitrary units. For spectra c to f the spacing between the dots is 500 kHz, for g the spacing is 250 kHz.

the triplet state, which has an average lifetime of 45 μ s (ref. 9), is ~ 40 μ s. In the counting interval of 10 s the molecule therefore spends a time long compared with the proton flip-flop time of ~ 30 μ s (ref. 10) in the non-magnetic singlet states.

The method reported here may be particularly useful for studying the properties of amorphous organic materials. The optical selection of a single molecule eliminates not only the static, inhomogeneous broadening of the probe molecules, but simultaneously any orientational anisotropy. This will allow magnetic-resonance experiments on random materials to yield information as if they were done on single-crystalline matrices. $\square$

Received 4 February; accepted 7 April 1993.

1. Sharnoff, M. *J. chem. Phys.* **46**, 3263–3264 (1967).
2. Kwiram, A. L. *Chem. Phys. Lett.* **1**, 272–275 (1967).
3. Schmidt, J. & van der Waals, J. H. *Chem. Phys. Lett.* **2**, 640–642 (1968).
4. van Dorp, W. G., Schaafsma, T. J., Soma, M. & van der Waals, J. H. *Chem. Phys. Lett.* **21**, 221–225 (1973).
5. Moerner, W. E. & Kador, L. *Phys. Rev. Lett.* **62**, 2535–2538 (1989).
6. Orrit, M. & Bernard, J. *Phys. Rev. Lett.* **65**, 2716–2719 (1990).
7. Ambrose, W. P., Basché, Th. & Moerner, W. E. *J. chem. Phys.* **95**, 7150–7163 (1991).
8. van Strien, A. J. & Schmidt, J. *Chem. Phys. Lett.* **70**, 513–517 (1980).
9. de Vries, H. & Wiersma, D. A. *J. chem. Phys.* **70**, 5807–5822 (1979).
10. Goldman, M. *Spin Temperature and Nuclear Magnetic Resonance in Solids* (Oxford Univ. Press, London, 1970).

ACKNOWLEDGEMENTS. One of us (J.K.) is supported in part by the Deutsche Forschungsgemeinschaft (DFG). We thank W. P. Ambrose and A. Krüger for crystal growth. This work is supported by the 'Stichting voor Fundamenteel Onderzoek der Materie FOM' with financial aid from the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek NWO'.

Optical detection of magnetic resonance in a single molecule

J. Wrachtrup*, C. von Borczyskowski*, J. Bernard†, M. Orrit† & R. Brown†

* Fachbereich Physik, Freie Universität Berlin, Arnimallee 14, 1000 Berlin 33, Germany

† Centre de Physique Moléculaire Optique et Hertzienne, u.a. 283 du C.N.R.S., Université Bordeaux I, 33405 Talence cedex, France

MAGNETIC resonance spectroscopy¹ is a powerful tool for molecular characterization and structure determination. The sensitivity of conventional approaches is limited to about 10^{10} electron spins or 10^{16} nuclear spins; this sensitivity can be improved to about 10^5 spins by polarizing the spins via optical pumping and detecting optical rather than microwave photons². Recently, fluorescence from single molecules was detected by tuning a single-frequency laser in the inhomogeneously broadened fluorescence excitation band of a dilute dispersion of pentacene in a host crystal of *p*-terphenyl^{3,4}. Here we report that, by combining single-molecule fluorescence spectroscopy with optically detected magnetic resonance for the pentacene-doped *p*-terphenyl system, we can detect magnetic resonance in a single pentacene molecule. We observe two of the three possible transitions between sublevels of the metastable triplet state. The spectral lineshapes indicate that the proton nuclear spin states change during the measurement, leading to spectral diffusion within the magnetic resonance line.

Single-molecule spectroscopy is usually done by observing fluorescence from a molecule driven several million times per second through the excitation-emission cycle. Occasionally, the molecule crosses over to the triplet manifold, and stays there for the average triplet lifetime. Consequently, fluorescence photons are emitted in packets separated by dark intervals when the molecule is in the triplet manifold⁵. As the different triplet sublevels generally have different population rates and different

lifetimes, irradiation at a resonant microwave frequency will affect the average fluorescence intensity⁶ by modifying the average duration of the dark intervals.

The optical apparatus used to isolate single molecule lines and the sample preparation are described in ref. 5. The setup is based on a single-mode dye laser which excites the sample through a single-mode, polarization-preserving optical fibre. The fluorescence is collected efficiently with a paraboloid mirror and sent to a photon-counting photomultiplier tube through a holographic filter and a red-pass glass. The microwaves are applied through a coaxial cable ending in a short-circuit loop, closely wrapped around the crystal in front of the mirror. All measurements were done in liquid helium at 1.8 K without application of an external magnetic field.

The experimental procedure was as follows. We first chose and recorded an isolated single molecule line in the red wing of the inhomogeneous O_1 line⁷, at $16,875\text{ cm}^{-1}$, 7 cm^{-1} off the maximum, tuned the laser frequency to resonance and saturated the optical transition by increasing the laser power. We then observed the change in fluorescence intensity while sweeping the microwave frequency. For correlation measurements, the photoelectron pulses were fed into a logarithmic correlator over accumulation periods of about 10 minutes.

Figure 1 shows the optically detected magnetic resonance (ODMR) of a single pentacene molecule in the region 1.3 to 1.5 GHz. We observe two ODMR lines which we attribute to transitions between the y and z ($y-z$) and between the x and z ($x-z$) triplet spin sublevels on the basis of data⁸ on pentacene ensembles. The z state is assumed to be negligibly populated and long-lived. The positions of the ODMR transitions are close to those for pentacene in *p*-terphenyl at room temperature⁹ and in naphthalene crystal at 1.2 K (ref. 8). Neither line was shifted by more than 3 MHz for the ~ 10 molecules in the red wing that we have investigated. The maximum intensity of the ODMR effect was $\sim 25\%$ for $x-z$ and 15% for $y-z$. Below the saturation laser power, the lineshape of both transitions is very asymmetrical (Fig. 1 inset). There is a steep rise (in about 200 kHz) on one side and a slow fall (over about 5 MHz for $x-z$, 3 MHz for $y-z$) on the other side. The microwave transitions became broader and more symmetrical on saturation.

The fact that the two transitions show up indicates that the x and y sublevels are comparably populated. The ODMR effect calculated for a single transition from known intersystem crossing rates^{5,10} would be 40%. This amount could be roughly shared in 25% on $x-z$ and 15% on $y-z$. The ODMR lineshapes and widths resemble those of an ensemble average for naphthalene

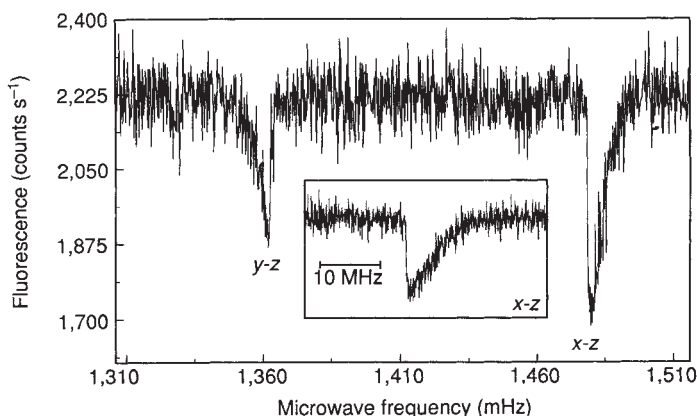


FIG. 1 Decrease of the fluorescence intensity of a single pentacene molecule when the microwave field is resonant with the $y-z$ or the $x-z$ transition. The inset shows the $x-z$ transition on an enlarged frequency scale, on reducing the microwave power by a factor of ~ 30 .

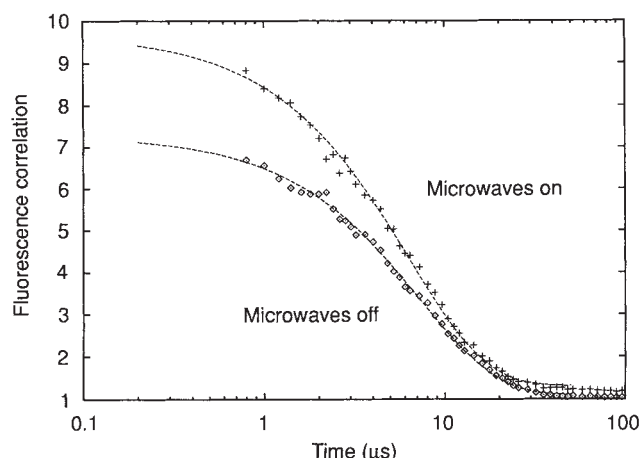


FIG. 2 Fluorescence autocorrelation function with the microwaves off or on resonance with the low-frequency side of the x - z transition, at a medium microwave power level. Smooth curves are mono-exponential fits described in the text.

in a durene crystal¹¹, which are determined mainly by second-order hyperfine interaction between the electronic spin and intramolecular proton spins.

The steep rise of the lines occurs over less than about 200 kHz. This value is similar to the homogeneous width deduced from electron spin echo experiments⁸. A single molecule should, at any given time, be in only one of its electron-nuclear spin states. But as several or all spin states show up, nuclear spins must change configuration during the detection time, leading to spectral diffusion within the ODMR lines. We do not know yet whether this diffusion arises in the singlet ground state or the triplet state, or is due to the optical excitation process itself.

We also measured the correlation function of the emitted photons. The results are plotted in Fig. 2 with excitation frequencies on and off resonance. The correlation decays are fairly well fitted by single exponentials⁵ (see Fig. 2). The contrast of the correlation clearly increases when the frequency is on resonance, whereas the decay time remains practically unchanged. The difference between the two experimental curves for times longer than 10 μ s (Fig. 2) is a reproducible effect and may be due to spin dynamics in the triplet state.

Although both x and y levels are populated, only a single exponential is seen in the correlation function (without microwave irradiation), which might point to comparable lifetimes for these states. The increase of the contrast when the microwaves are turned on is expected: longer dark intervals mean stronger correlation of photon bunches. The correlation decay rate is roughly given by the bunch duration⁵ (under conditions of optical saturation) and, as is observed, should hardly depend on the average triplet lifetime.

Single-molecule ODMR extends the high sensitivity and selectivity of single-molecule spectroscopy to the general field of magnetic resonance¹. Techniques such as electron-nuclear double resonance, optically induced nuclear spin polarization^{12,13} and ODMR in external magnetic fields⁶ can be envisaged immediately. Studying the influence of the local environment on fine and hyperfine structures, magnetic resonance spectroscopy of trace compounds and orientation-dependent investigations in external magnetic fields of single molecules in disordered matrices will eventually become feasible. Further, the correlation method can now be applied to microscopic processes affecting spin states. $\square$

4. Ambrose, W. P. & Moerner, W. E. *Nature* **349**, 225-227 (1991).
5. Bernard, J., Fleury, L., Talon, H. & Orrit, M. *J. chem. Phys.* **98**, 850-859 (1993).
6. Clark, R. H. (ed.) *Triplet State ODMR Spectroscopy* (Wiley, New York, 1982).
7. Ambrose, W. P., Basché, Th. & Moerner, W. E. *J. chem. Phys.* **95**, 7150-7163 (1991).
8. van Strien, A. J. & Schmidt, J. *Chem. Phys. Lett.* **70**, 513-517 (1980).
9. Sloop, D., Liu, H. L., Lin, T. S. & Weissman, S. I. *J. chem. Phys.* **75**, 3746-3757 (1981).
10. de Vries, H. & Wiersma, D. A. *J. chem. Phys.* **70**, 5807-5822 (1979).
11. Hutchison, C. A. Jr, Nicholas, J. V. & Scott, G. W. *J. chem. Phys.* **53**, 1906-1917 (1970).
12. Stehlik, D. in *Excited States* (ed. Lim, E. C.) Vol. 3, 204 (Academic, New York, 1977).
13. von Borczyskowski, C. in *Advances in Magnetic Resonance* (ed. Waugh, J.) Vol. 12, 113 (Academic, New York, 1988).

ACKNOWLEDGEMENTS. This work was supported by the DFG (Sfb 337) and a joint Procope grant from the Ministère des Affaires Étrangères and the DAAD.

Increased chlorine dioxide over Antarctica caused by volcanic aerosols from Mount Pinatubo

S. Solomon*, R. W. Sanders*, R. R. Garcia† & J. G. Keys‡

* Aeronomy Laboratory, NOAA, Boulder, Colorado 80303, USA

† National Center for Atmospheric Research, Boulder, Colorado 80307, USA

‡ National Institute for Water and Air Research, Lauder, Central Otago, New Zealand

THE annual springtime depletion of Antarctic ozone¹ has been shown to be due to the action of chlorine species, activated by reactions occurring on the surfaces of polar stratospheric clouds (PSCs)^{1,2}. Similar reactions may also take place on the surfaces of liquid sulphuric acid aerosols when the temperature is too high to permit the formation of PSCs³⁻⁴. Such processes may have been facilitated following the eruption of Mount Pinatubo in June 1991, when unprecedented amounts of sulphur compounds were injected into the stratosphere⁵. Here we present observations of Antarctic chlorine dioxide abundances in the austral autumn and winter of 1991 (when aerosol concentrations were at background levels) and 1992 (greatly enhanced aerosol concentrations). We find that in 1992, unlike 1991, chlorine dioxide levels increased dramatically in the autumn, when PSCs were extremely unlikely to have been present. Model results suggest that this was mainly caused by the direct activation of chlorine nitrate on the aerosol surfaces. The effect of the Pinatubo aerosols probably contributed to the unprecedented depth and areal extent of Antarctic ozone depletion in 1992.

Aerosols from Mount Pinatubo were transported to the Antarctic stratosphere mainly after October 1991, following the spring break-up of the polar vortex⁶. A period of particular interest is the polar autumn of the following year (March to April 1992) before polar stratospheric clouds are likely to be present, allowing study of any chemical changes due to sulphuric acid aerosols alone.

Observations of chlorine dioxide (OCIO) provide a sensitive test for enhanced chlorine radical chemistry and associated ozone destruction⁷⁻¹⁰. Since February 1991, we have made spectroscopic measurements of chlorine dioxide at McMurdo Station in Antarctica, using the same instrumentation and data analysis procedures as were used in the Antarctic spring (September-October) of 1986 and 1987. The chlorine dioxide absorption features near 408-409 and 420-421 nm are derived from spectra of scattered light from the sky using nonlinear least-squares analysis methods¹⁰. Analysis of the spectra as a function of slant path through the atmosphere showed no detectable chlorine dioxide in early March in 1991 or 1992. Spectra from these periods were therefore used as background to analyse observations made later in the autumn season. The analysis was applied to measurements obtained in polar autumn of 1991 from 16 March to 10 May 1991. No chlorine dioxide could be detected at any time during the austral autumn.

Received 3 March; accepted 7 April 1993.

1. Abragam, A. *The Principles of Nuclear Magnetism* (Oxford Univ. Press, London, 1961).
2. Suter, D. *J. Magn. Reson.* **99**, 495-506 (1992).
3. Orrit, M. & Bernard, J. *Phys. Rev. Lett.* **65**, 2716-2719 (1990).

The seasonal evolution of chlorine dioxide in the autumn of 1992 was markedly different from that of 1991. Figure 1a shows representative measurements. The first day when chlorine dioxide could be clearly detected was 13 April 1992 (day number 104). It was observed on every subsequent day until the sky became too dim to allow further measurements (8 May). There was a systematic diurnal pattern with values being higher in the afternoon than in the morning, as shown in more detail in Fig. 1b for two additional days. This figure also shows the contrast between April measurements in 1991 and 1992.

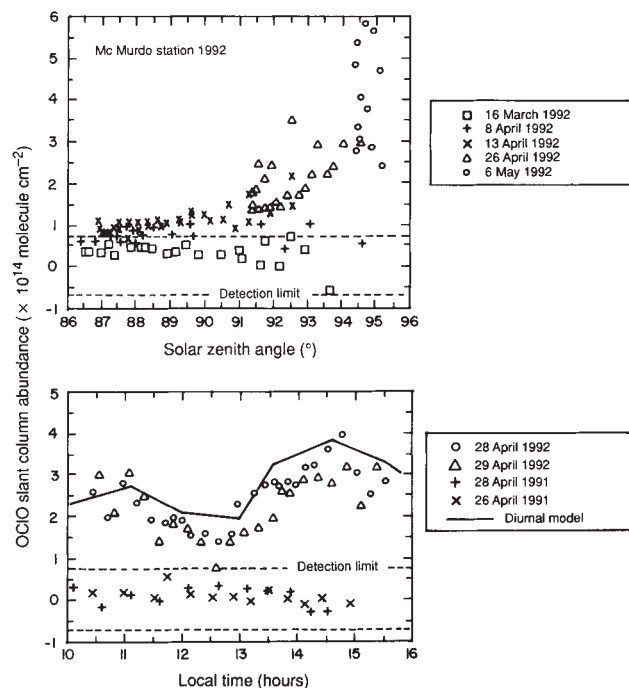
In the absence of heterogeneous chemistry, stratospheric reactive chlorine is expected to reside largely in the reservoir species, HCl and ClONO₂, and slant column abundances of chlorine dioxide about 100 times smaller than those measured in autumn 1992 (Fig. 1) are predicted. The observed enhancements in chlorine dioxide above these values depend on release of chlorine radical species from one or both of these reservoirs as a result of surface reactions, and on the abundance of reactive nitrogen remaining in the stratosphere (which can reconvert chlorine radicals to ClONO₂). We studied these factors using a detailed diurnal model⁷ in which the amount of total reservoir chlorine and the fraction released by surface reactions are prescribed over an assumed altitude range from 18 to 24 km. Figure 1b shows the calculated slant column variation obtained when it is assumed that 30% of the reservoir chlorine has been released from 18 to 24 km and that no reactive nitrogen is present. Similar values could be obtained by adopting different fractional reservoir chlorine release over different ranges or altitudes, or assuming small amounts of reactive nitrogen to be present. A range of calculations indicates that the observed chlorine dioxide abundances are consistent with release of ~30–50% of the chlorine residing in reservoir species for assumed reactive nitrogen abundances of 0–100 parts per 10¹² by volume (p.p.t.v.) from 15 to 25 km. The calculated diurnal variation agrees with that observed, displaying a minimum at local noon and afternoon values that are about $1\text{--}2 \times 10^{14}$ molecule cm⁻² greater than those obtained in the morning near 92–93° solar zenith angles. The calculated afternoon increase results from the gradual photolysis of the Cl₂O₂ formed especially at night as a function of time. The photolysis is slow at the large solar zenith angles typical of this time of year at McMurdo (Fig. 1).

These observations are strong evidence for release of reactive chlorine radical species from reservoir compounds during the

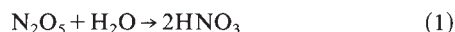
autumn season in 1992 but not in 1991, suggesting a likely role for surface chemistry. It is clear that Pinatubo aerosols were present in large amounts in the Antarctic stratosphere by autumn 1992 (but not in 1991), but it is also critical to examine the likelihood of polar stratospheric clouds being present then. Arctic^{11,12} and Antarctic¹³ observations suggest that initial formation of polar stratospheric clouds often requires supersaturated conditions, and confirm that polar stratospheric clouds first begin to form near 30–50 mbar at threshold temperatures below 195 K, much colder than the local temperatures at McMurdo in April (Fig. 2). On some occasions, clouds have not been observed in non-denitrified air even at temperatures below 193 K (ref. 12). The Microwave Sounding Unit (MSU) satellite observations provide information on the lowest temperatures obtained anywhere in the Antarctic vortex over the region from 12–24 km. Those measurements show that 1992 was no colder than 1991 over this averaged height range, and display minimum values during 15–30 April 1992 between 202 and 206 K, too warm for polar stratospheric cloud formation (A. O'Neill, personal communication). Meteorological analyses from the European Centre for Medium-range Weather Forecasting (ECMWF) show minimum temperatures of 208 K at 50 mbar on 13 April 1992 (the first day of observation of OCIO shown in Fig. 1). Temperatures earlier in the season are warmer, and are within 1 K of the 1991 minimum temperatures from 1 to 21 April. Even at the end of April, ECMWF minimum temperatures were no lower than 199 K at 50 mbar. Isolated patches of colder temperatures too small to be resolved either by the satellite or the meteorological analysis cannot be ruled out, but it is difficult to see how a small cloud could produce the persistently higher OCIO levels observed daily after 13 April at McMurdo. Localized cooling events would also require supersaturation before cloud formation would be expected¹². Finally, satellite and ground-based lidar methods have been used to measure polar stratospheric clouds in Antarctica^{13,14}. Available measurements reveal that they first form in June in most years and only occasionally in late May, so cloud formation in April would appear to be unusual and would require temperatures colder than normal, which is not indicated by the 1992 temperature data.

We next consider in more detail the chemical perturbations that the liquid volcanic particles may produce. Laboratory studies^{2,3,15} show that the following reactions take place in liquid

FIG. 1a, Slant column abundances of chlorine dioxide observed as a function of solar zenith angle on representative days during autumn 1992 using visible spectroscopy. The spectrograph is a crossed Czerny–Turner system with a Reticon diode array detector cooled to about -80°C (refs 7–10). Spectra were obtained with the twilight sky as the light source, pointing at an 80° angle toward the solar direction at twilight⁹. The background spectrum was obtained at a solar zenith angle of $\sim 92^\circ$ on the morning of 9 March 1992; other background spectra obtained in March gave essentially the same results. No chlorine dioxide was measurable in March and early April. April 13 is the first day when measurable chlorine dioxide was present. By April 26, very large abundances were obtained, with larger values measured in the afternoon than in the morning (see text and Fig. 2). b, Observations of chlorine dioxide on representative days in late April of 1991 and 1992. In 1991, the chlorine dioxide was below the detection limit throughout the day, whereas in 1992 much larger values were readily measured. A systematic diurnal pattern was obtained on all days in late April 1992, with a minimum near local noon (13:00 hours at McMurdo Station), and larger values in the afternoon than in the morning. A diurnal model calculation, where it is assumed that 30% of the chlorine residing in reservoir species has been released, is shown for comparison⁷. The total reservoir chlorine¹⁸ (scaled to 1992 values) and the fraction released by surface reactions are prescribed over an assumed altitude range from 18 to 24 km. Total reactive bromine abundances of 10 p.p.t.v. are assumed²⁰. The photochemical reactions of chlorine, bromine and nitrogen species are considered in detail as a function of solar zenith angle, as well as the slant path through the atmosphere relevant to the twilight measurement geometry.



sulphuric acid particles:



Both processes form HNO_3 , thereby suppressing the abundance of nitrogen dioxide and indirectly enhancing reactive chlorine radicals. In addition, reaction (2) also releases HOCl , which photolyses rapidly to produce ClO . The ClO radical can go on to destroy ozone and, through reaction with BrO , is believed to be the chemical source of the chlorine dioxide observed. Reaction (1) proceeds rapidly on all sulphuric acid surfaces regardless of temperature, whereas reaction (2) is a strong function of the percentage of water in the liquid particle and hence is expected to proceed most rapidly under cold stratospheric conditions. A two-dimensional dynamical-chemical model was used to examine the expected impact of these reactions, adopting Antarctic satellite extinction observations to prescribe the relevant aerosol surface areas (L. R. Poole, personal communication). These measurements show maximum surface areas near 18–20 km in February 1992 of $\sim 20\text{--}25 \mu\text{m}^2 \text{cm}^{-3}$, about a factor

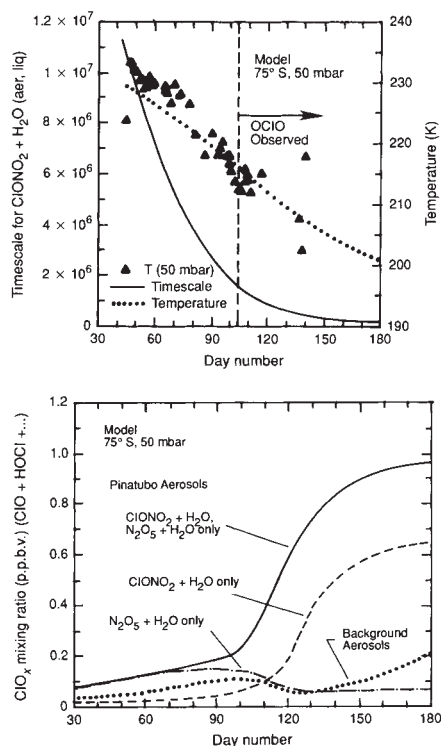


FIG. 2 a, Timescale for uptake of ClONO_2 on sulphuric acid aerosols in the Antarctic vortex during autumn, 1992 (solid line; times are in seconds) together with temperatures from the two-dimensional dynamical/chemical model²¹ (dotted line) used in this study (which are in excellent agreement with McMurdo balloonsonde data during 1991, indicated by the triangles). A temperature-dependent reaction probability of $0.006 \times e^{-0.15 \times (T-200)}$ has been assumed, which is based on laboratory studies of this surface reaction². b, Calculated abundance of reactive chlorine species (defined as $\text{HOCl} + \text{ClO} + \text{Cl} + 2 \times \text{Cl}_2\text{O}_2 + \text{OCIO}$) from the two-dimensional model of Garcia and Solomon²¹ for various assumptions regarding surface area and chemistry of sulphuric acid particles. Satellite observations of aerosol optical depths in Antarctica were used to infer the surface areas present there under clean conditions and in the presence of the aerosols from Mount Pinatubo (L. R. Poole, personal communication). Bromine chemistry was not modelled in detail but it was assumed that ~ 7 p.p.t.v. of BrO was present above 16 km during the day, as indicated by *in situ* measurements in Antarctica²⁰. The solid line shows the behaviour expected for 1992, assuming a distribution of Pinatubo aerosols based on satellite measurements and including reactions of N_2O_5 and ClONO_2 with H_2O on these surfaces. Much lower amounts of reactive chlorine are predicted when only background aerosols are considered together with reactions (1) and (2). Calculations are also shown for cases where only N_2O_5 or only ClONO_2 was assumed to react with the Pinatubo aerosols.

of 20 greater than those for unperturbed conditions. Figure 2 illustrates the 50 mbar temperatures obtained in the model and in McMurdo balloonsondes as a function of day number near 78°S during the autumn season, along with the associated timescale for destruction of ClONO_2 by reaction (2). There is a rapid decrease in the timescale for uptake of ClONO_2 on the aerosols as the Antarctic stratosphere cools from warm summer temperatures to autumn and winter values¹⁶, falling from about 4 months in early February (day 40) to about 2 weeks by day 104, when chlorine dioxide was first observed. Thus by day 104 it is reasonable to expect that reactive chlorine will be rapidly released from the chlorine nitrate reservoir by reactions in liquid sulphuric acid particles. The build-up of this released reactive chlorine depends critically on the amount of NO_2 present, which controls the reformation of ClONO_2 . This in turn is strongly coupled to photolysis of the HNO_3 reservoir. As discussed previously¹⁷, the photolysis rate of HNO_3 is therefore a key factor in determining how effectively chlorine radicals destroy ozone once they are released. The rate of photolysis of HNO_3 becomes slow at the large solar zenith angles characterizing the Antarctic autumn stratosphere, implying that NO_2 abundances will decrease rapidly and any chlorine radicals released will be less and less effectively reconverted to ClONO_2 as the autumn season progresses.

Figure 2 shows a set of four model calculations that reveal the expected seasonal variation of stratospheric reactive chlorine (defined here as $\text{HOCl} + \text{ClO} + 2 \times \text{Cl}_2\text{O}_2 + \text{Cl} + \text{OCIO}$) at 50 mbar (20 km) for various aerosol conditions. When only background aerosols are considered together with reactions (1) and (2), the reactive chlorine maximizes near day 100 and reaches about 100 p.p.t.v. total (with ClO representing ~ 70 p.p.t.v.). Although these values represent important changes from gas-phase chemistry that are likely to result in some ozone loss, they are not sufficient to produce measurable chlorine dioxide during the fall. When Pinatubo aerosols are considered along with reactions (1) and (2), reactive chlorine abundances increase dramatically near day 110, close to the point when our observations first reveal chlorine dioxide (Fig. 1). Figure 2 shows that reaction (1) plays a role in determining when high values of reactive chlorine are achieved (mainly through its importance in suppressing NO_2), but is not sufficient on its own greatly to enhance the reactive chlorine abundances. At the large solar zenith angles characterizing this time of year (Fig. 1), little of the reactive chlorine resides in the form of ClO , making OCIO an excellent observable for detecting this chemistry.

These findings imply that considerable chemical destruction of ozone occurred in the 1992 Antarctic stratosphere at latitudes, altitudes and times of the year when the air was too warm for polar stratospheric clouds to be present, contributing to the large and extensive⁶ ozone losses observed in Antarctica during 1992 (when the ozone hole was $\sim 15\text{--}25\%$ larger than in previous years; P. A. Newman, personal communication, and World Meteorological Organization, 1992). The ozone depletion is expected to vary strongly with latitude, because of the temperature dependence of reaction (2) and, more importantly, the zenith angle dependence of HNO_3 photolysis. Considerable ozone changes are predicted for southern mid-latitudes assuming roughly the same vertical profile of volcanically enhanced aerosol surface (as indicated by observations). The calculations suggest substantial ozone losses by November 1992 near 20 km due to volcanic aerosols of $\sim 15\%$ and 10% , at 61°S and 54°S , respectively. Larger ozone losses could be obtained if air parcel excursions increased the exposure of the chemically perturbed air to sunlight¹⁸. The quantitative details of the calculations presented are sensitive to factors such as the aerosol surface area prescribed and the rate constants for chemical processes (for example, any temperature dependence in HNO_3 photolysis rate¹⁹, and important uncertainties in the surface chemistry on liquid and frozen sulphuric acid particles). Nonetheless, our results indicate that reactions on sulphuric acid aerosols should

be considered in attempts to understand ozone trends in polar regions and at mid-latitudes. Although similar effects are also expected in the Arctic, they are likely to be more difficult to detect against the larger variability of Arctic ozone concentrations, making Antarctica a unique natural laboratory for studying this chemistry. □

Received 10 December 1992; accepted 24 March 1993.

1. Solomon, S., Garcia, R. R., Rowland, F. S. & Wuebbles, D. J. *Nature* **321**, 755–758 (1986).
2. *World Meteorological Organization, Rep. No. 25* (1992).
3. Tolbert, M. A., Rossi, M. J. & Golden, D. M. *Geophys. Res. Lett.* **15**, 851 (1988).
4. Brasseur, G., Granier, C. & Walters, S. *Nature* **348**, 626 (1990).
5. Bluth, G. J. et al. *Geophys. Res. Lett.* **19**, 151–154 (1992).
6. Hofmann, D. J. et al. *Nature* **359**, 283 (1992).
7. Solomon, S., Sanders, R. W. & Miller, H. L. Jr *J. geophys. Res.* **95**, 13807 (1990).
8. Solomon, S., Mount, G. H., Sanders, R. W. & Schmeltekopf, A. L. *J. geophys. Res.* **92**, 8329 (1987).
9. Sanders, R. W. et al. *J. geophys. Res.* (in the press).
10. Solomon, S., Mount, G. H., Sanders, R. W., Jakoubek, R. O. & Schmeltekopf, A. L. *Science* **242**, 550–555 (1988).
11. Arnold, F., Petzoldt, K. & Reimer, E. *Geophys. Res. Lett.* **19**, 677–680 (1992).
12. Dye, J. E. et al. *J. geophys. Res.* **97**, 8015–8034 (1992).
13. Collins, R. L., Bowman, K. P. & Gardner, C. S. *J. geophys. Res.* **98**, 1001–1010 (1993).
14. McCormick, M. P., Trepte, C. R. & Pitts, M. C. *J. geophys. Res.* **94**, 11241–11252 (1989).
15. Hanson, D. R. & Ravishankara, A. R. *J. geophys. Res.* **96**, 5081–5090 (1991).
16. Turco, R. P. & Hamill, P. *Ber. Bunsenges phys. Chem.* **96**, 323–334 (1992).
17. Hofmann, D. J. & Solomon, S. *J. geophys. Res.* **94**, 5029 (1989).
18. Jones, R. L. et al. *Geophys. Res.* **94**, 11529 (1989).
19. Rattigan, O., Lutman, E. R., Jones, R. L. & Cox, R. A. *Ber. Bunsenges phys. Chem.* **96**, 399–404 (1992).
20. Brune, W. H., Anderson, J. G. & Chan, K. R. *J. geophys. Res.* **94**, 16639 (1989).
21. Solomon, S., Garcia, R. R. & Stordal, F. *J. geophys. Res.* **90**, 12981–12989 (1985).

ACKNOWLEDGEMENTS. We thank A. O'Neill and L. Poole for helpful comments and for allowing us to use MSU and SAGE data. We also appreciate comments on the manuscript and discussions with A. R. Ravishankara, David Hofmann, O. B. Toon and A. Tuck.

Elevated consumption of carbon relative to nitrogen in the surface ocean

Raymond N. Sambrotto*, Graham Savidge†, Carol Robinson‡, Philip Boyd†, Taro Takahashi*, David M. Karl§, Chris Langdon*, David Chipman*, John Marra* & Louis Codispoti||

* Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

† Queen's University of Belfast, School of Biology and Biochemistry, Marine Biology Station, Portaferry, County Down, BT22 1PF, UK

‡ University of Wales, Bangor, School of Ocean Sciences, Menai Bridge, Gwynedd LL59 5EY, UK

§ SOEST, University of Hawaii, Honolulu, Hawaii, 96822, USA

|| Monterey Bay Aquarium Research Institute, 160 Central Avenue, Pacific Grove, California 93950, USA

UPTAKE of atmospheric CO₂ by the ocean's 'biological pump' is driven by export of carbon from the euphotic zone to deeper waters^{1,2}. As nitrate is a limiting nutrient in large regions of the ocean, measurements of nitrate uptake are often used to estimate the amount of carbon exported in this way^{3–6}. This presupposes knowledge of the molar C:N ratio in the organic material exported from the upper waters, which is usually taken to be 6.6 (the Redfield ratio^{7,8}). Recent studies have suggested, however, that the consumption ratio of C:N may deviate from this value in coastal waters^{9–11}. Here we present time-series from both coastal waters and open-ocean sites which demonstrate that net organic carbon production greatly exceeded that predicted from nitrate consumption and the Redfield C:N ratio. We found a similar discrepancy in sections across broad regions of the North Atlantic during eutrophic periods. Our results suggest that extrapolating from nitrate consumption using the Redfield ratio leads to significant underestimates of organic carbon export from the euphotic zone.

Conversions such as that from nitrate consumption to organic carbon production are crucial both to account for the ocean's

role in the global carbon cycle and to relate the net production of organic matter (based on O₂ budgets), new production and carbon flux quantitatively^{12–15}. There have, however, been few direct measurements of the effect of plankton productivity on the inorganic carbon system of the surface ocean. Isolated studies have indicated that the consumption of dissolved inorganic carbon (DIC) relative to NO₃ may considerably exceed the Redfield C/N ratio¹⁶. We have now recorded examples in both coastal and oceanic regions from which time-series of the DIC and NO₃ changes during eutrophic periods could be constructed^{9–11,17} (Fig. 1).

To estimate biological consumption from the observed local changes, we must consider effects not directly associated with the production of organic matter: concentration changes due to evaporation and precipitation; advection and diffusion; the production of calcium carbonate; and the exchange of gaseous CO₂ with the atmosphere, we addressed evaporation and precipitation by normalizing all values to a constant salinity appropriate for each region and time. The physics of the continental shelf regions (southeast Bering Sea and Gerlache Strait) were characterized and the effects of advection evaluated. In these regions, the measured flow fields and observed gradients indicate that the raised carbon to nitrogen ratios in Fig. 1 were not due to advection. The movement of eddies through a region can alter the local surface water chemistry over short periods¹⁸, but there was no evidence of such advective bias in the May 1989 sampling. The effect of CaCO₃ production can be estimated from changes in alkalinity. On this basis, there was no effect in Gerlache Strait, and in the North Atlantic and the Bering Sea time series, less than 10% and ~31% respectively of the total DIC loss could be attributed to CaCO₃ production. During the periods in Fig. 1, biological productivity was typically >1 g C m⁻² d⁻¹, probably indicating the influx of atmospheric CO₂ (ref. 19). Table 1 summarizes these aspects of the DIC system for the regions in Fig. 1 and indicates that using the Redfield ratio to convert new (nitrate) production to carbon underestimates the actual net organic carbon production by more than 50% on average.

Sections in the North Atlantic also suggested that the net consumption of carbon to nitrate during productive periods may be significantly greater than Redfield stoichiometry (Fig. 2). In the eastern basin, meridional differences in pre-formed nutrients complicate the analysis. Therefore, we estimated the regional

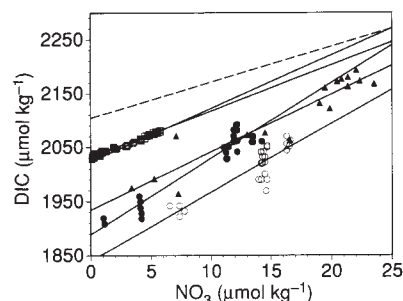


FIG. 1 Relationships between surface-layer dissolved inorganic carbon (DIC) and NO₃ during time series at four locations. Linear regression analyses of the salinity-normalized data are shown. For each region the salinities, the resulting slopes, their 95% confidence interval, and the probability that the slopes could be the Redfield value (6.6) are: ○, Southeast Bering mid-shelf (57.5° N, 165° W, May 1980) [33.00 parts per thousand, p.p.t.], 12.67 ± 3.05, $p < 0.001$; ●, Southeast Bering mid-shelf (May 1981) [33.00 p.p.t.], 14.00 ± 1.66, $p < 0.001$; ▲, Gerlache Strait (64° S, 61.5° W, December–January 1986–87) [33.40 p.p.t.], 10.68 ± 2.08, $p = 0.01$; □, North Atlantic (47° N, 20° W, 20 April–8 May 1989) [35.50 p.p.t.], 9.88 ± 0.69, $p < 0.001$; ■, (47° N 20° W, 18 May–31 May 1989) [35.50 p.p.t.], 8.54 ± 1.63, $p = 0.05$. Data from the surface isothermal layer were used except in the case of Gerlache Strait in which all data from 0–40 m were used. Mixed layers with NO₃ concentrations less than 1 μM were excluded from the Bering Sea data to avoid possible nutrient limitation effects. The dashed line indicates the Redfield slope (6.6).

TABLE 1 Comparison of the two net community production estimates

Region	Measured ΔNO_3	Measured ΔDIC (% organic)	Atmosphere influx*	NCP† (DIC budget)	NCP ($\Delta\text{NO}_3 \times$ Redfield C/N)
Southeast Bering Sea (May 1980)	10	125 (69%) ⁹	13 (ref. 27)	99	66
Southeast Bering Sea (May 1981)	11.5	150 (69%) ⁹	15 (ref. 27)	119	76
Gerlache Strait (November 15–January 15)	19	210 (100%) ¹⁰	16	226	125
47° N, 20° W (April 1989)	3	30 (90%) ¹⁷	4	31	20
47° N, 20° W (May 1989)	2	20 (90%) ¹⁷	NA	> 18	13.2

Estimates based on (1) budget of DIC system measurements, and (2) the conversion of new (NO_3) production by the Redfield C/N ratio. This comparison is based on the surface mixed-layer changes and both approaches neglect the vertical supply of material and thus are similarly conservative. The passage of some of the atmospheric influx to deeper waters is also ignored, but is likely to be small. All values are concentrations (μM).

* Based on air–sea difference in partial pressure of CO_2 and wind speed, to estimate the areal flux of CO_2 ($\text{mol m}^{-2} \text{d}^{-1}$). This was normalized to the surface-mixed-layer depth to obtain concentration units.

† Organic fraction of ΔDIC plus atmospheric flux.

relationship between mixed-layer DIC and nitrate prior to spring phytoplankton growth. The marked change from the pre-formed relationship observed in May was similar to those found in the time series and corresponded to a period associated with widespread phytoplankton growth in this region²⁰. Sections across Georges Bank (a productive shelf region) also recorded a DIC/ NO_3 relationship well above the Redfield ratio (Fig. 2), and this difference persisted after accounting for the effects of water mass mixing¹¹. The decrease in the surface DIC/ NO_3 ratio between May and June in the eastern North Atlantic may have been due both to the invasion of atmospheric CO_2 and to the respiration of organic carbon as the NO_3 levels approached depletion, a condition that was widespread by June. Also, the coverage of the oceanic sections ($\sim 1,400 \text{ km}$) was larger than the diameter of typical eddies in this region (100–200 km; ref. 18) and is likely to have averaged large-scale changes accurately. Thus, the onset of phytoplankton growth in the North Atlantic during spring has a much larger impact on the surface-water DIC system than would be expected from the changes in NO_3 concentrations and the Redfield ratio.

Two questions are raised by this non-Redfield behaviour: what mechanism(s) bring(s) about the observed ratios, and how are the observed consumption ratios to be reconciled (as they must be) with thermocline stoichiometry? Although several processes could contribute to elevated DIC/ NO_3 consumption

ratios, not all are viable explanations for the present observations. For example, new nitrogen can be supplied by nitrogen fixation and atmospheric deposition. The cyanobacteria that generate substantial N_2 fixation in tropical regions²¹ were not, however, a significant component of the plankton in the mid- and high-latitude regions discussed here. A similar discrepancy was observed between the DIC/ PO_4 consumption ratio and the Redfield value, and this cannot be due to N_2 fixation. Atmospheric deposition of nitrogen is probably small in comparison with pre-formed nutrients during the spring bloom for regions far from sources of industrial pollution²². Also, it is difficult to justify a vertical diffusive flux of nutrients that is enriched in nitrate relative to DIC to account for the elevated DIC/ NO_3 ratios observed. Relative vertical flux must reflect surface consumption, and in both the Gerlache Strait and Bering Sea studies, the inclusion of data from deeper water further increased the discrepancy between the net consumption ratios and the Redfield ratio.

The excess loss of DIC from surface waters is best explained by *in situ* biological processes that recycle nitrogen more efficiently than carbon. This implies that the organic material exported from surface waters is enriched in carbon. In most of the studies cited here, the C:N ratio in the surface layer particulate material was similar to or less than the Redfield ratio (for example, southeast Bering Sea, 6.5; 47° N, 20° W, 5.5). Such particulate ratios are typical for phytoplankton growth in nutrient-rich environments and remained low even in the June section from the northeast Atlantic in which low nutrient ($< 2 \mu\text{M NO}_3$) conditions were encountered. However, the average C:N ratio of the material caught in sediment traps at 100 m in the Bransfield Strait study was 8.82 (weighted by C flux) and the loss of such carbon-rich material explains some (but not all) of the elevated DIC/ NO_3 consumption ratio observed in the overlying water. Additional carbon may be consumed in the production of dissolved organic matter (DOM) that is relatively nitrogen-poor. The levels of DOM measured in Gerlache Strait cannot account for the excess DIC consumed, although this is the only study from which DOM data are available. The process by which the elevated DIC/ NO_3 surface consumption ratios are reconciled with thermocline stoichiometry also may be closely associated with ocean biology. For example, surface-dwelling, upper-trophic-level consumers in the ocean's food web may respire much of the excess organic carbon produced by the plankton back to the atmosphere before it reaches the thermocline²³. Alternatively, the re-equilibration of the consumption ratios with thermocline stoichiometry may be brought about by decomposition processes as the organic matter descends. In either case, the longer timescales associated with the annual cycle or the movement of the surface signal to

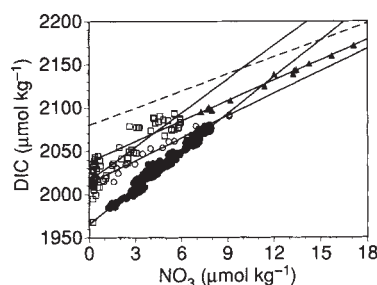


FIG. 2 Relationships between mixed-layer DIC and NO_3 in sections from the North Atlantic. Pre-formed nutrient concentrations along 20° W ($\blacktriangle$) were based on measured values²⁸ together with extrapolations from subsurface (150–400 m) water from 47° N to 60° N²⁹. These data were normalized to a salinity of 35.5 units and produced a slope of 7.85 ± 0.51 . All other data are from surface-mixed layer measurements. The slopes of the latter data, their 95% confidence intervals, and the probabilities that the slopes are not different from the slope of the pre-formed line (or from 6.6 for Georges Bank) are as follows: $\bullet$, 20° W section sampled 19 May–24 May 1989, [34.50 p.p.t.], 14.16 ± 0.20 , $p < 0.001$; $\circ$, same section sampled 1 June–6 June 1989, [35.20 p.p.t.], 8.60 ± 0.71 , $p < 0.05$; $\square$, Georges Bank frontal region¹¹, [32.90 p.p.t.], 13.02 ± 1.62 , $p < 0.001$. The dashed line indicates the Redfield slope (6.6).

depth may ameliorate much of the discrepancy between thermocline stoichiometry and the short-timescale surface changes that we report.

Our observations indicate that carbon cycling is more complex than has been appreciated because the consumption of carbon relative to limiting nutrients from the surface ocean can differ significantly from the ratio at which these materials are supplied. This negates the use of Redfield ratios to translate from production, in terms of limiting nutrients, to the carbon system, despite earlier hopes that various measures of ocean productivity could be reconciled in this manner⁶. But as smaller amounts of nitrate are required to generate a given net production signal, previous discrepancies found between O₂ utilization rates (OUR) and nitrate upwelling¹² as well as between O₂ budgets and local productivity²⁴ are reduced. Also, similar to the OUR estimates of biological production, budgets based on the surface DIC system are less prone to sampling bias than are instantaneous rate measurements. A second implication is that the ocean's biological pump encompasses an additional degree of freedom that was not part of the original conceptual model¹. The measured consumption ratios suggest that carbon export exceeds 15–20% of total production that was estimated from the nitrogen *f*-ratio (the ratio of new to total nitrogen production³) and is probably closer to 25–35% of total production. Coupled with recent estimates of ocean production (~64 gigatonnes C yr⁻¹; ref. 25), carbon export from surface ocean waters may exceed 15 gigatonnes C yr⁻¹. This is twice as large as export estimates from sediment-trap measurements²⁶, and suggests that the growth of ocean plankton has been underestimated as a sink for atmospheric CO₂. □

Received 2 November 1992; accepted 1 April 1993.

1. Volk, T. & Hoffert, M. I. in *The Carbon Cycle and Atmospheric CO₂ Natural Variations Archaean to Present* (Sundquist, E. T. & Broecker, W. S.) 99–110 (AGU, Washington, 1985).
2. Sarmiento, J. L. & Sundquist, E. T. *Nature* **356**, 589–593 (1992).
3. Eppley, R. W. & Peterson, B. J. *Nature* **282**, 677–680 (1979).
4. Eppley, R. W., Renger, E. H. & Betzer, P. R. *Deep-Sea Res.* **30**, 311–323 (1983).
5. Chavez, F. P. & Barber, R. T. *Deep-Sea Res.* **34**, 1220–1243 (1987).
6. Platt, T. *et al. Mar. Ecol. Progr. Ser.* **52**, 77–88 (1989).
7. Redfield, S. C., Ketchum, B. H. & Richards, F. A. in *The Sea* (ed. Hill, M. N.) 26–77 (Wiley, New York, 1963).
8. Takahashi, T., Broecker, W. S. & Langer, S. J. *geophys. Res.* **90**, 6907–6924 (1985).
9. Codispoti, L. A., Friederich, G. E. & Hood, D. W. *Continental Shelf Res.* **5**, 133–160 (1986).
10. Karl, D. M., Tilbrook, B. D. & Tien, G. *Deep-Sea Res.* **38**, 1097–1126 (1991).
11. Sambrotto, R. N. & Langdon, C. *Continental Shelf Res.* (in the press).
12. Jenkins, W. J. & Goldman, J. C. *J. mar. Res.* **43**, 465–491 (1985).
13. Jenkins, W. J. *Nature* **331**, 521–523 (1988).
14. Emerson, S., Quay, P., Stump, C., Wilbur, D. & Knox, M. *Global biogeochem. Cycles* **5**, 49–69 (1991).
15. Minas, H. J. *ICES Marine Sci. Symp. Series* (eds Li, W. K. W. & Maestrini, S. Y.) (International Council for the Exploration of the Sea, Copenhagen, in the press).
16. Cooper, L. H. N. *J. mar. biol. Assoc. UK* **18**, 677–728 (1933).
17. Chipman, D., Marra, J. & Takahashi, T. *Deep-Sea Res.* **40**, 151–169 (1993).
18. Savidge, G. *et al. Progr. Oceanogr.* **29**, 235–281 (1992).
19. Watson, A. J., Robinson, C., Robertson, J. E., Williams, P. J. le B. & Fasham, M. J. R. *Nature* **350**, 50–53 (1991).
20. Esaias, W. E., Feldman, G. C., McClain, C. R. & Elrod, J. A. *Eos* **67**, 835–837 (1986).
21. Carpenter, E. J. & Romans, K. *Science* **254**, 1356–1358 (1991).
22. Duce, R. A. in *The Role of Air–Sea Exchange in Geochemical Cycling* (ed. Baut-Menard, P.) 497–530 (Reidel, Dordrecht, 1986).
23. Huntley, M. E., Lopez, M. D. G. & Karl, D. M. *Science* **253**, 64 (1991).
24. Schulenberger, E. & Reid, J. L. *Deep-Sea Res.* **28**, 901–919 (1981).
25. Karl, D. M. *Int. Geosphere-Biosphere Progr. Symp. Global Change Tokyo, April, 1992* (in the press).
26. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. W. *Deep-Sea Res.* **34**, 267–285 (1987).
27. Hood, D. W. & Codispoti, L. A. in *The Potential Effects of Carbon Dioxide Induced Climatic Change in Alaska* (ed. McBeath, J. H.) 33–39 (Univ. of Alaska, 1984).
28. Sambrotto, R. N., Martin, J. H., Broenkow, W. W., Carlson, C. & Fitzwater, S. E. *Deep-Sea Res.* **40**, 441–457 (1993).
29. Glover, D. M. & Brewer, P. G. *Deep-Sea Res.* **35**, 1525–1546 (1988).

ACKNOWLEDGEMENTS. We thank P. J. le B. Williams for comments on the developing manuscript, and J. R. Toggweiler for constructive comments. This work was supported in part by the U.S. Division of Ocean Sciences.

How geometrical constraints contribute to the weakness of mature faults

David A. Lockner & James D. Byerlee

U.S. Geological Survey, 345 Middlefield Road, Menlo Park, California 94025, USA

INCREASING evidence that the San Andreas fault has low shear strength¹ has fuelled considerable discussion regarding the role of fluid pressure in controlling fault strength. Byerlee^{2,3} and Rice⁴ have shown how fluid pressure gradients within a fault zone can produce a fault with low strength while avoiding hydraulic fracture of the surrounding rock due to excessive fluid pressure. It may not be widely realised, however, that the same analysis^{2–4} shows that even in the absence of fluids, the presence of a relatively soft 'gouge' layer surrounded by harder country rock can also reduce the effective shear strength of the fault. As shown most recently by Byerlee and Savage⁵, as the shear stress across a fault increases, the stress state within the fault zone evolves to a limiting condition in which the maximum shear stress within the fault zone is parallel to the fault, which then slips with a lower apparent coefficient of friction than the same material unconstrained by the fault. Here we confirm the importance of fault geometry in determining the apparent weakness of fault zones, by showing that the apparent friction on a sawcut granite surface can be predicted from the friction measured in intact rock, given only the geometrical constraints introduced by the fault surfaces. This link between the sliding friction of faults and the internal friction of intact rock suggests a new approach to understanding the microphysical processes that underlie friction in brittle materials.

Byerlee and Savage⁵ analysed a fault containing a layer of granular material (fault gouge) that behaves as a Coulomb material; in other words, it deforms elastically up to a failure

limit, and then plastically with continued deformation. They were able to explain, on theoretical grounds (as Rice⁴ had shown for a ductile 'von Mises' layer), the rotation of the maximum compressive stress in the fault zone to an angle of 45° to the fault—a phenomenon that previously had only been assumed⁶, or observed empirically⁷. The Byerlee–Savage model has the following important consequences.

(1) It provides an explanation for the development of R₁ and R₂ Riedel shears (Fig. 1)⁶ commonly observed in both natural and artificial fault zones^{8–10}. These features coincide with the Coulomb yield surfaces that occur at angles of $\pm(\pi/4 - \phi/2)$ to the direction of maximum principal stress, σ_1 .

(2) The incompatibility between the orientations of the Coulomb yield surfaces and the direction of average shear strain within the gouge suggests that the bulk deformation will be accommodated locally by cooperative slip on conjugate surfaces oriented in the directions of the Coulomb yield surfaces.

(3) Yielding on the Coulomb surfaces is assumed to occur at

$$\mu_f = \tan \phi = \tau_1 / (\sigma_1 + \sigma_c) \quad (1)$$

where ϕ is the friction angle, τ_1 and σ_1 are stresses on the local yield surfaces within the gouge, and σ_c is tensile (cohesive) strength (Fig. 1b). However, the apparent coefficient of friction μ_a resolved parallel to the fault surfaces is given by

$$\mu_a = \tan \chi = \tau / (\sigma + \sigma_c) \quad (2)$$

where χ can be considered the apparent friction angle. μ_a is the coefficient of friction that is reported in standard laboratory deformation experiments.

(4) It can be shown by simple construction (Fig. 1b) that $\tan \chi = \sin \phi$ and therefore

$$\mu_a = \sin(\tan^{-1} \mu_f) \quad (3)$$

One consequence of equation (3) is that the apparent coefficient of friction of a Coulomb material deformed in simple shear between competent fault blocks will always be less than the true

coefficient of friction. This systematic error has been recognized for some time in soil mechanics¹¹, where shear box tests are found to give consistently lower values of shear strength than triaxial tests using bulk samples. The Byerlee-Savage model provides a theoretical framework for understanding the measurement bias inherent in the simple shear test.

(5) A further consequence of equation (3) is that for the simple shear geometry, the apparent coefficient of friction will in fact be limited to $\mu_a \leq 1$, regardless of the magnitude of the true coefficient of friction. This upper bound to μ_a is consistent with experimentally derived friction values. For example, Byerlee¹² found that for a broad range of rock types and for normal stresses up to 1,700 MPa, sliding friction was limited to $0.55 \leq \mu_a \leq 0.90$, a relation often referred to as Byerlee's law.

The model concludes that μ_a provides a lower bound to the intrinsic coefficient of friction of a Coulomb material. This is the result of geometric effects related to the presence of the fault surfaces. If these are removed from the system, it may be possible to measure μ_f directly. We assert that a logical upper bound for μ_f can be obtained by testing a sample of intact rock of the same composition as the fault gouge under consideration. In this case, the interlocking and welding of grains can only increase the strength of the material above the frictional strength. By measuring the peak strength of the intact rock at the time of

failure, we obtain a measure of the resistance to shearing in the absence of a pre-existing fault. To test this hypothesis, we combine old strength data for Westerly granite^{13,14} with the additional new measurements in Fig. 2. The open squares are measurements of internal friction μ_i taken from fracture experiments on intact samples. Internal friction is normally defined as the derivative of the failure envelope in shear-normal stress space. However, to be consistent with the preceding analysis, we will use a different but easy to measure parameter:

$$\mu_i = \tau_p / (\sigma_p + \sigma_c) \quad (4)$$

where τ_p and σ_p are the peak shear and normal tractions resolved on the eventual fracture surface. (Effects resulting from curvature in the failure envelope are neglected in this analysis.) The solid diamonds in Fig. 2 are direct measurements of μ_a from triaxial sawcut experiments. The open circles are predicted values of μ_a obtained by applying equation (3) to the observed μ_i data. The agreement between observed and predicted values of μ_a is remarkable. Apparently, for this example, μ_i provides not only an upper bound but is in fact equal to μ_f . This exercise shows how large (especially at lower normal stress) the discrepancy between μ_f and μ_a can be.

In general, additional characteristics of the intact rock, such as porosity and grain cementation, must be included in the relation between μ_i and μ_f . For example, Dunn *et al.*¹⁵ reported a power-law relation between strength and porosity in sandstones over a porosity range of 2 to 30%. More porous samples also tend to have lower internal friction, even though the sliding friction μ_a should be about equal in all cases. Thus we expect that a porosity correction similar to that reported by Dunn *et al.* must be included in a complete expression relating μ_i and μ_f . At present, this would be little more than empirical curve fitting, and is not included in our analysis. As the porosity of an actively deforming gouge¹⁶ ranges between 8 and 14%, it is not obvious why the intact granite, with 1% porosity at the time of failure, should agree so well with the theory. It may be because the zone of intense microcrack damage which develops where the shear fracture first forms is much more intensely fractured¹⁷, and therefore more porous, than the surrounding intact rock. The internal friction data shown in Fig. 2 can then be interpreted as representing the Coulomb properties of this damage zone at the time that fracture initiates. Direct observations of the changes in local stresses and porosity are difficult to obtain and will require additional precise testing.

The comparison shown in Fig. 2 provides a link between internal friction and friction on pre-existing faults. This is satisfying, as the processes responsible for the brittle fracture strength and frictional sliding strength are essentially the same. Friction is a phenomenological term defined by equation (2) or some

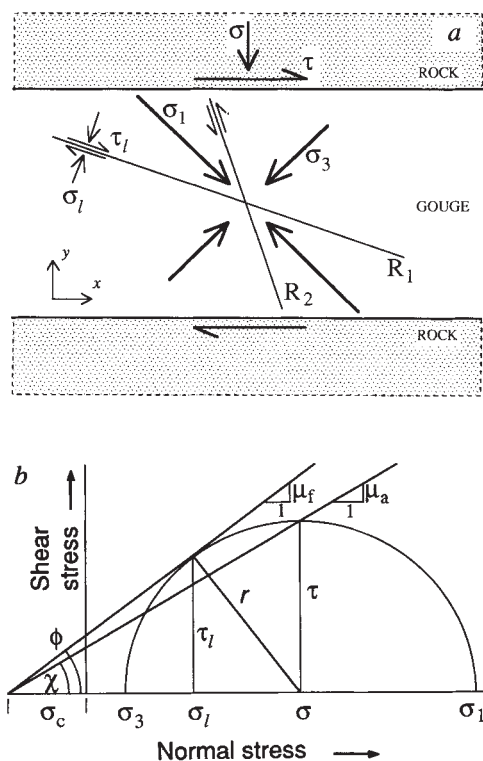


FIG. 1 *a*, Geometry of fault containing Coulomb gouge used in Byerlee-Savage model. Consistent with the observation of Mandl *et al.*⁷, the model predicts that for a well-developed fault, the maximum principal stress σ_1 within the gouge rotates to 45° to the fault surfaces. Then R_1 and R_2 Riedel shears, which are observed in many faults, correspond to the Coulomb failure planes within the gouge. *b*, Mohr diagram, plotting shear stress against normal stress, showing stress state for mature fault zone. ϕ and χ are, respectively, the true and apparent friction angles for the fault system. $\mu_f = \tan \phi = \tau_l / (\sigma_l + \sigma_c)$ represents the local yield condition on oblique failure surfaces within the gouge. Rigid boundary constraints impose a stress state in which $\sigma_{xx} = \sigma_{yy} = \sigma$, a condition that can only be satisfied when σ_1 is oriented 45° to the fault. The strength of the fault zone is generally represented by the boundary traction τ and σ giving an apparent coefficient of friction $\mu_a = \tan \chi = \tau / (\sigma + \sigma_c)$. As $|\tau| = |\tau_l|$, we have $\sin \phi = r / (\sigma + \sigma_c) = \tan \chi$, which limits the strength of this fault geometry to $\mu_a \leq 1$.

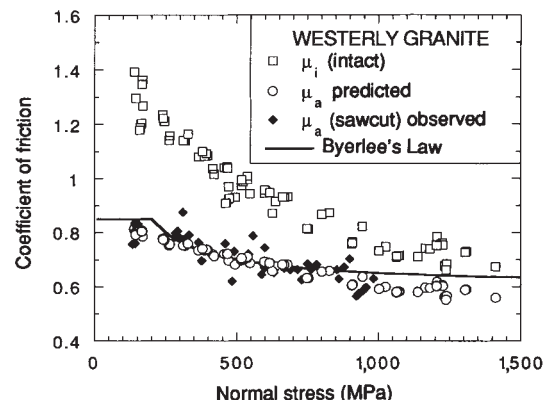


FIG. 2 Application of equation (3) to internal friction data ($\square$) successfully predicts the observed sawcut friction values ($\blacklozenge$) and Byerlee's law.

similar relation; it does not tell us the physical mechanisms responsible for the observed shear strength. If we restrict ourselves to brittle deformation, then micromechanical mechanisms of grain rotation, grain crushing and crack growth are what result in the observed shear strength of a fault. These are the same processes that lead to brittle failure in an intact rock. The principal difference between the two cases is whether or not a fault is present and the geometrical constraints that the fault introduces. Viewed in this way, the result shown in Fig. 2 should not be surprising.

To extend this result further, consider the convergence of μ_i and μ_a in Fig. 2 at high normal stress. Byerlee¹⁸ noted this more than 20 years ago, identifying it with the brittle-ductile transition in rock, and it has since been noted for a variety of rock types¹⁹. Increasing normal stress tends to suppress the dilatancy and open porosity that is required for individual grains within a gouge to roll past each other. It thus leads to more dense, interlocked grains which must deform by fracture and micro-crack growth. With increasing normal stress, gouge looks and behaves more and more like intact rock, until a point is reached where it becomes as easy to break intact grains as it is to shear on existing fault surfaces. At this point (the brittle-ductile transition) we expect a convergence of μ_i and μ_a as well as a transition from localized to more distributed shear. This transition should occur whether or not crystal plasticity is taking place on the microscopic scale.

A key element of the Byerlee-Savage model is the zero-strain boundary condition

$$e_{xx} = e_{zz} = 0 \text{ (within fault gouge)} \quad (5)$$

imposed by the competent wall rock on the gouge layer. A consequence of this is that as the gouge deforms it must evolve towards a stress state in which $\sigma_{xx} = \sigma_{yy}$, a condition which can only occur when σ_1 is oriented at 45° to the fault. This boundary condition is what distinguishes the mechanics of fault deformation from deformation of material in bulk, it may imply that all mature faults develop this internal stress state, regardless of the details of the gouge constitutive law. For example, Rice⁴ analysed a shear zone containing a ductile von Mises material subject to equation (5), and obtained the same stress orientation as for the Byerlee-Savage Coulomb gouge. The plastic yielding inherent in this model also tends to suppress the hydraulic fracturing that can accompany a build-up of pore pressure p within the fault zone. A necessary condition for hydraulic fracturing is $p > \sigma_3$. With increasing p , however, effective stress is reduced, causing the gouge to weaken and eventually yield. Because of the geometric constraints on the gouge³, the plastic yielding increases σ_3 so that it remains above p .

The approach presented here represents a departure from the traditional view of friction based on ductility and plastic yielding in metals. In that view, the true area of contact between two surfaces increases with normal stress because of plastic yielding; shear strength, which is proportional to the true contact area, also increases, making the coefficient of friction relatively insensitive to normal stress. By contrast in our conceptual model the true coefficient of friction, represented by a function involving μ_i , can be relatively sensitive to normal stress. This sensitivity is then suppressed through the $\sin \phi$ dependence imposed by the simple shear constraints of the fault zone geometry. Although plastic yielding of contact spots may be an appropriate model for friction in metals, it provides a poor representation of a fault in brittle, crustal rock. In this case, shearing on fault surfaces rapidly produces a layer of gouge which is more appropriately analysed as discussed here. This effect may explain why the steady-state velocity dependence in recent state-variable constitutive laws for friction appears to be negative for bare surfaces and yet positive for mature faults with gouge^{16,20}. Stability analysis of such models²¹ has shown that negative velocity dependence is necessary for fault instability. Thus it may be necessary to modify the constitutive model to include additional

effects (such as fracture of asperities, fault healing and sealing, fault interactions, and fluid pressure changes) to explain earthquake instabilities. □

Received 17 December 1992; accepted 6 April 1993.

- Hickman, S. H. *Rev. Geophys. IUGG Rep.* 759-775 (1991).
- Byerlee, J. *Geophys. Res. Lett.* **17**, 2109-2112 (1990).
- Byerlee, J. D. *Tectonophysics* **211**, 295-303 (1992).
- Rice, J. R. in *Fault Mechanics and Transport Properties of Rocks* (eds Evans, B. & Wong, T.-f.) 475-503 (Academic, London, 1992).
- Byerlee, J. D. & Savage, J. C. *Geophys. Res. Lett.* **19**, 2341-2344 (1992).
- Hansen, B. in *Pro. 5th int. Conf. Soil Mech. Found. Engng* 127-131 (Dunod, Paris, 1961).
- Mandl, G., de Jong, L. N. J. & Maltha, A. *Rock Mechanics* **9**, 95-166 (1977).
- Logan, J. M., Friedman, M., Higgs, N., Dengo, C. & Shimamoto, T. *U.S. geol. Surv. Open File Rep.* 79-1239, 305-343 (1979).
- Moore, D. E., Summers, R. & Byerlee, J. D. *J. struct. Geol.* **11**, 329-342 (1989).
- Moore, D. E. & Byerlee, J. D. *Tectonophysics* **211**, 305-316 (1992).
- Saada, A. S., Fries, G. & Ker, C.-C. *Soils Found.* **23**, 98-112 (1983).
- Byerlee, J. D. *Pure appl. Geophys.* **116**, 615-626 (1978).
- Byerlee, J. D. *J. geophys. Res.* **72**, 3639-3648 (1967).
- Summers, R. & Byerlee, J. *U.S. geol. Surv. Open File Rep.* 77-142, 129 (1977).
- Dunn, D. E., LaFountain, L. & Jackson, R. *J. geophys. Res.* **78**, 2403-2417 (1973).
- Marone, C., Raleigh, C. B. & Scholz, C. H. *J. geophys. Res.* **95**, 7007-7025 (1990).
- Lockner, D. A., Moore, D. E. & Reches, Z. in *33rd U.S. Rock Mechanics Symp.* (eds Tillerson, J. R. & Wawersik, W. R.) 807-816 (Balkema, Rotterdam, 1992).
- Byerlee, J. D. *J. geophys. Res.* **73**, 4741-4750 (1968).
- Ismail, I. A. H. & Murrell, S. A. F. *Tectonophysics* **175**, 237-248 (1990).
- Beeler, N. M., Weeks, J. D. & Tullis, T. E. *Eos* **73**, 511 (1992).
- Rice, J. R. & Ruina, A. L. *J. appl. Mech.* **50**, 343-349 (1983).

Mass-spectrometric U-series dates for Israeli Neanderthal/early modern hominid sites

F. McDermott*†, R. Grün‡, C. B. Stringer§ & C. J. Hawkesworth*

* Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK

‡ Quaternary Dating Research Centre, Australia National University, GPO Box 4, Canberra 2601, Australia

§ Department of Palaeontology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK

THE nature of the relationship between Neanderthals and early modern *Homo sapiens* is controversial, yet it is fundamental to our understanding of early human evolution^{1,2}. The Middle Palaeolithic sites of Israel are critical to this debate, because unlike those of western Europe and Africa they contain both Neanderthal (at Tabun³ and Kebara⁴ for example) and anatomically modern hominids (as at Skhul⁵ and Qafzeh⁶). Here we present new mass spectrometric ²³⁰Th/²³⁴U dates for dental fragments from the Middle Palaeolithic burial sites of Tabun, Qafzeh and Skhul. These data, combined with published ages from electron spin resonance (ESR), provide compelling evidence that the Tabun Neanderthals and Qafzeh early modern *Homo sapiens* were approximately coeval in the southern Levant some 100 ± 5 kyr ago, but indicate that some of the Skhul material is younger. The study also shows that combined mass-spectrometric ²³⁰Th/²³⁴U and ESR dating is an invaluable technique for dating archaeological sites beyond the range of radiocarbon dating.

Uncertainties in the chronology of Tabun, Qafzeh, Skhul and other key sites have given rise to conflicting views of the Neanderthal/early modern relationship in the Levant. In the early 1980s the Tabun Neanderthals were generally considered to be 50-60 kyr old, and they were apparently succeeded by *Homo sapiens* at ~40 kyr⁷ (for example, at Qafzeh and Skhul), thereby allowing the possibility that Neanderthals had contributed to the ancestry of modern humans. This view was reinforced when

† Present address: Department of Geology, University College Dublin, Belfield, Dublin 4, Ireland.

the layers containing the Kebara Neanderthal skeleton were dated at 60 ± 4 kyr by thermoluminescence⁸ and at 62 ± 8 kyr by ESR⁹. But subsequent thermoluminescence dating of Qafzeh flints yielded ages of 92 ± 5 kyr^{10,11} in agreement with the biostratigraphy¹². Moreover, the Skhul and Qafzeh sites yielded ESR dates in the 80–120 kyr range^{13,14}, and similar or even older ages were obtained for the Tabun Neanderthal-bearing layers^{15,16}. These thermoluminescence and ESR dates are still viewed with scepticism by many palaeoanthropologists^{2,17,18} and a key objective of this study is to exploit the enhanced sensitivity and improved precision offered by mass-spectrometric U-series dating in order critically to evaluate the published ESR dates. U-series dates for fossil teeth reflect the time elapsed since they acquired, through groundwater, their inventory of uranium following death and burial, and so they tend to be minimum ages^{19,20}, although combined U-series and ESR data can overcome this problem²¹. Here, we present new high-precision (± 1 –2%) mass-spectrometric $^{230}\text{Th}/^{234}\text{U}$ dates for small (~ 50 –100 mg) aliquots of mechanically separated dentine (DE)

and enamel (EN) from bovid tooth samples that had been analysed previously by ESR^{13–16}. In practice, more than 70% of the samples studied here yield concordant U-series and early-uptake ESR ages, which implies a closed system behaviour²¹. Furthermore, where we have obtained ages for dentine and enamel from the same tooth (samples 551 and 854; Table 1), minor discrepancies only were detected, despite large differences in U concentrations. This indicates that the U uptake history was similar for the enamel and dentine.

Detailed descriptions of the three Israeli sites studied have been given elsewhere^{3,6,22–24}. The Tabun site is important because of the occurrence of well-documented Neanderthal fossils attributed to layer C (Fig. 1a), and because its archaeological and stratigraphic sequence has been used extensively for Late Pleistocene correlation throughout the Levant^{3,25}. But the chronology of Tabun is extremely controversial (Fig. 1b). In Jelinek's chronology⁷ (column 1, Fig. 1b) layers B to C date from about 40 to 60 kyr, layer D from about 65–80 kyr, and layer E from 80–100 kyr. An alternative chronology has been proposed on the basis of the archaeological sequence, microfaunal constraints and absolute dates from other sites²⁵ (column 2, Fig. 1b). In the second scheme, a relatively long hiatus is inferred between layers C and D, and so layer D is placed at about 100 kyr with layer E at 110–150 kyr. Significantly, both schemes place layer C and the associated Neanderthals in the 50–60 kyr time interval, and they contrast sharply with that based on ESR dates¹⁵ (column 3, Fig. 1b). In the latter scheme layer C, and by implication the Tabun Neanderthal fossils, is ~ 110 kyr old, with the upper part of layer E (Ea) dated at 150–170 kyr.

In this study, two dentine samples from Tabun layer Ea yield precise U-series ages of 159.1 ± 1.3 and 168.1 ± 2.6 kyr, respectively, indistinguishable from the relatively imprecise early uptake ESR dates¹⁵ of 158 ± 41 and 167 ± 42 kyr (Table 1). Similarly, the U-series age of 110.7 ± 0.9 kyr for sample 556EN (layer D) supports the chronology in ref. 15 (Fig. 1b). Moreover, three samples from the Neanderthal-bearing layer C (551DE and 551EN and 522DE) yield new U-series dates of 97.8 ± 0.4 , 101.7 ± 1.4 and 105.4 ± 2.6 kyr respectively, and all are within error of the early uptake ESR ages¹⁵ (Fig. 1b). Thus, the U-series dates offer compelling new evidence for the antiquity of the Tabun site, and make previously published chronologies^{7,25} untenable.

Qafzeh cave (lower Galilee) has yielded the remains of at least 20 hominids^{6,26} which are classified as *Homo sapiens*, albeit with some archaic cranial features^{6,27}. Archaeological dating schemes had indicated an age of < 50 kyr^{3,7,28}, although amino-acid analyses of bone²⁹ and microfaunal constraints suggested older ages, possibly 70–100 kyr. More recently, thermoluminescence dates on burnt flint⁸ and early uptake ESR dates on mammal teeth¹⁴ yielded average ages of 92 ± 5 kyr and 100 ± 10 kyr, respectively. In this study, sample 368DE yielded a U-series age of 106.4 ± 2.4 kyr, identical to its early uptake ESR age (105 ± 2 kyr)¹⁴. Similarly, the U-series age for sample 371EN is 88.6 ± 3.2 , within error of its average early uptake ESR age (103 ± 19 kyr)¹⁴. Thus, the new U-series ages for Qafzeh layer XIX are in the range 85–110 kyr, which corroborates the thermoluminescence and early uptake ESR dates, and strongly supports the contention that the Qafzeh *Homo sapiens* are significantly older than the Kebara Neanderthal, and may be coeval with those of Tabun layer C. As in the Tabun site, the new U-series dates for Qafzeh are consistent only with early uptake ESR dates and not with the linear uptake ESR dates favoured in some ESR studies^{13–16} (Fig. 2).

The small cave of Es Skhul is located near Tabun^{5,13}. Layer B yielded cranial and post-cranial remains of at least 10 hominids representing an archaic type of modern *Homo sapiens* and showing skeletal affinities with those from Qafzeh^{6,27,30}. Two well preserved bovid teeth from layer B had been dated previously by the ESR technique and yielded early uptake ages of 88.1 ± 13.1 and 68.0 ± 5.4 for samples 521 and 522, respectively¹³.

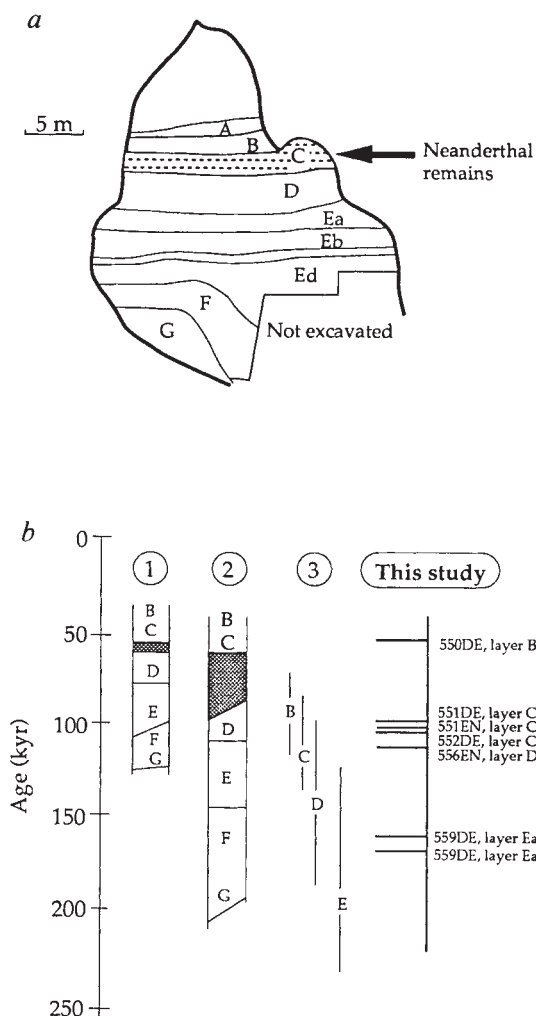


FIG. 1 a, Simplified stratigraphic section through the Tabun cave site after ref. 22. The Neanderthal remains are generally attributed to layer C (arrow). A more detailed stratigraphic section is provided in ref. 3. b, Comparison of the published chronologies of Tabun with the new U-series dates presented here. Column 1 is after Jelinek⁷, column 2 summarizes Bar-Yosef's scheme²⁵ and column 3 represents the early uptake (EU) ESR dates in ref. 15. Shaded areas represent inferred breaks in the stratigraphy. The U-series results from this study corroborate the EU ESR ages of ref. 15 and imply that the Tabun Neanderthals are 100 ± 5 kyr old, or almost twice as old as previously supposed.

TABLE 1 New mass spectrometric U-series data for dental fragments from Tabun, Qafzeh and Skhul

Sample	^{238}U ($\mu\text{g g}^{-1}$)	^{230}Th (pg g^{-1})	$(^{230}\text{Th}/^{232}\text{Th})$	$(^{234}\text{U}/^{238}\text{U})$	$(^{230}\text{Th}/^{238}\text{U})$	U-series age (kyr)	EU-ESR (kyr)	LU-ESR (kyr)
Tabun								
Layer B								
550DE	9.6591 ± 13	66.568 ± 0.170	346.75 ± 1.87	1.0902 ± 07	0.4085 ± 11	$50.69 \pm^{0.23}_{0.23}$	76 ± 14	85 ± 18
Layer C								
552DE	1.0153 ± 05	10.906 ± 0.150	$2,475.00 \pm 35.0$	1.0230 ± 05	0.6367 ± 88	$105.36 \pm^{2.58}_{2.52}$	111 ± 30	113 ± 31
551EN	0.3388 ± 01	3.692 ± 0.016	192.15 ± 2.66	1.0559 ± 34	0.6459 ± 29	$101.69 \pm^{1.36}_{1.34}$	121 ± 29	134 ± 36
551DE	5.3366 ± 34	58.526 ± 0.098	96.81 ± 0.46	1.0848 ± 10	0.6501 ± 11	$97.84 \pm^{0.43}_{0.42}$	121 ± 29	134 ± 36
Layer D								
556EN	11.4621 ± 30	137.09 ± 0.210	$4,184.00 \pm 7.00$	1.0963 ± 30	0.7089 ± 11	$110.68 \pm^{0.88}_{0.87}$	93 ± 12	152 ± 24
Layer Ea								
559DE{1}	7.7121 ± 12	104.418 ± 0.035	$1,399.00 \pm 3.00$	1.0372 ± 03	0.8026 ± 27	$159.05 \pm^{1.33}_{1.31}$	158 ± 41	158 ± 56
559DE{2}	5.9964 ± 09	82.546 ± 0.034	329.67 ± 1.54	1.0307 ± 20	0.8160 ± 34	$168.10 \pm^{2.61}_{2.53}$	167 ± 42	196 ± 57
Qafzeh								
Layer XIX								
371EN	0.0675 ± 08	0.7834 ± 0.010	20.26 ± 2.98	1.2098 ± 46	0.6879 ± 136	$88.61 \pm^{3.24}_{3.12}$	103 ± 19	125 ± 22
368DE	1.6530 ± 13	21.4240 ± 0.260	42.34 ± 0.83	1.2031 ± 18	0.7682 ± 93	$106.35 \pm^{2.36}_{2.31}$	105 ± 02	115 ± 08
Skhul								
Layer B								
521DE	7.6949 ± 12	73.360 ± 0.280	298.826 ± 1.14	1.0746 ± 08	0.5651 ± 22	$80.27 \pm^{0.55}_{0.55}$	88 ± 13	102 ± 18
522EN	3.6995 ± 06	21.336 ± 0.069	369.590 ± 1.48	1.0957 ± 12	0.3419 ± 11	$40.43 \pm^{0.21}_{0.21}$	68 ± 05	98 ± 11
854DE	7.9236 ± 13	45.360 ± 0.031	375.940 ± 0.11	1.0678 ± 09	0.3393 ± 23	$41.41 \pm^{0.39}_{0.38}$	55 ± 05	65 ± 05
854EN	0.4815 ± 50	2.831 ± 0.015	64.420 ± 0.84	1.0629 ± 34	0.3485 ± 19	$43.03 \pm^{0.47}_{0.46}$	55 ± 05	65 ± 05
856DE{1}	37.9845 ± 57	231.930 ± 0.440	665.990 ± 2.57	1.0932 ± 07	0.3619 ± 07	$43.46 \pm^{0.14}_{0.14}$	46 ± 05	66 ± 05
856DE{2}	37.9823 ± 45	228.090 ± 1.460	$2,993.80 \pm 15.00$	1.0382 ± 65	0.3560 ± 23	$45.53 \pm^{0.74}_{0.73}$	46 ± 05	66 ± 05

EN, enamel; DE, dentine. ^{238}U and ^{230}Th concentrations were determined by isotope dilution using ^{235}U and ^{229}Th tracers. $^{234}\text{U}/^{238}\text{U}$ ratios were measured in static mode with ^{234}U on the ion-counting channel of a multicollector mass-spectrometer. Chemical separation procedures for U and Th are based on those in ref. 31. Total blanks were $<2 \times 10^{10}$ atoms of ^{238}U and ^{232}Th , $<1 \times 10^6$ atoms of ^{230}Th and $<5 \times 10^6$ atoms of ^{234}U . Measured atomic ratios were converted to activity ratios using $\lambda_{230} = 9.195 \times 10^{-6} \text{ yr}^{-1}$, $\lambda_{238} = 1.551 \times 10^{-10} \text{ yr}^{-1}$ and $\lambda_{234} = 2.835 \times 10^{-6} \text{ yr}^{-1}$. Errors on the activity ratios are $\pm 2\sigma$, based on the within-run counting statistics. Weighing errors, estimated at 0.15%, have been taken into account when calculating the errors on ^{238}U and ^{230}Th concentrations. Duplicate analyses (559DE and 856DE) were carried out on different portions of the dentine. Also shown for comparison are tooth enamel ESR ages¹¹⁻¹³.

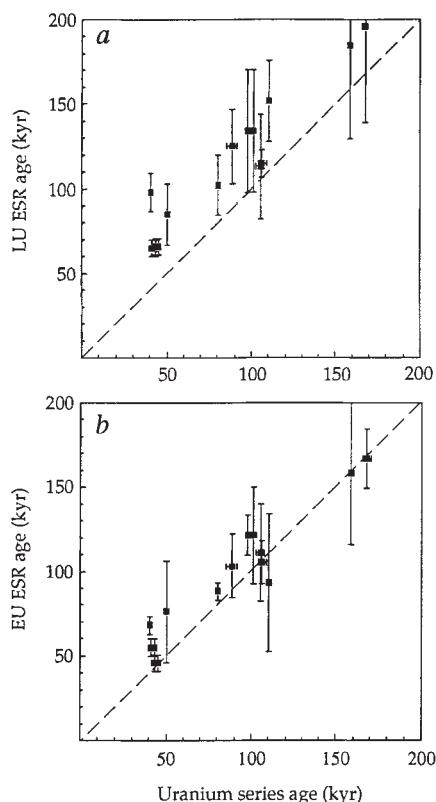


FIG. 2 *a*, Linear uptake (LU) ESR ages plotted against $^{230}\text{Th}/^{234}\text{U}$ ages for the same sample aliquots, showing that the LU ESR dates are probably too old. *b*, Early uptake (EU) ESR ages are broadly similar to the U-series ages and are within error for $>70\%$ of the samples. Error bars for the U-series ages are smaller than the data symbols except where shown.

In this study, a $^{230}\text{Th}/^{234}\text{U}$ age of 80.3 ± 0.6 has been obtained for sample 521DE, but sample 522EN yields a relatively young age of 40.4 ± 0.2 kyr indicative of post-depositional U uptake. Two *Dicerorhinus* teeth from layer B (samples 854 and 856) were also dated, and these yield rather uniform $^{230}\text{Th}/^{234}\text{U}$ dates in the range 41.4 ± 0.4 to 45.5 ± 0.7 kyr. Significantly, the dentine and enamel fractions of sample 854 yield similar ages (41.4 ± 0.4 versus 43.0 ± 0.5 kyr), despite large differences in their U concentration (Table 1). Thus, the U-series determinations combined with the published ESR data suggest a greater complexity in the stratigraphy of Skhul than was previously assumed, with clear evidence for at least two faunal ages within layer B. The younger closed system ages for samples 854 and 856 may be consistent with the suggestion that the Skhul hominids fall into earlier (Skhul 3, 6–10) and later (1, 4–5) assemblages²⁴, but further research will be required to investigate this possibility.

In summary, this first application of mass-spectrometric U-series dating to dental material provides strong support for the revised, more ancient chronologies for the Tabun and Qafzeh sites based on early uptake ESR age estimates. The excellent correlation between the early uptake ESR and U-series dates for more than 70% of the samples studied enhances considerably our confidence in the ages reported. Beyond the specific problems of the Middle Palaeolithic sequence of the Middle East, the feasibility of mass-spectrometric U-series dating of small dental fragments has been demonstrated, illustrating its potential for widespread further application. In detail, the data indicate that the mode of U uptake by dental materials is site-specific, and that the use of generalized U uptake models is hazardous. Finally, the conclusion that early modern *Homo sapiens* were probably coeval with and locally pre-dated Neanderthals in the Levant now seems inescapable, which in turn requires that the simple ancestor/descendant evolution models be abandoned. □

Received 25 January; accepted 10 March 1993.

1. Stringer, C. B. & Grün, R. *Nature* **351**, 701–702 (1991).
2. Wolpoff, M. H., Frayer, D. W. & Stringer, C. B. *Nature* **356**, 200–201 (1992).
3. Jelinek, A. J. *Science* **216**, 1369–1375 (1982).

4. Rak, Y. *Am. J. phys. Anthropol.* **81**, 323–332 (1990).
5. McCown, T. D. & Keith, A. *The Stone Age of Mount Carmel* Vol. II (Clarendon, Oxford, 1939).
6. Vandermeersch, B. *Les Hommes Fossiles de Qafzeh, Israel* (CNRS, Paris, 1981).
7. Jelinek, A. J. in *The Transition from Lower to Middle Palaeolithic and the Origin of Modern Man* (ed. Ronen, A.) 57–101 (Brit. Archaeol. Rep. Int. Ser. 151, Oxford, 1982).
8. Valladas, H. *et al. Nature* **330**, 159–160 (1987).
9. Schwarcz, H. P. *et al. J. archaeol. Sci.* **16**, 653–659 (1989).
10. Valladas, H. *et al. Nature* **331**, 614–616 (1988).
11. Aitken, M. J. & Valladas, H. *Phil. Trans. R. Soc. Lond. B* **337**, 139–144 (1992).
12. Bar-Yosef, O. & Vandermeersch, B. in *Préhistoire du Levant* (eds Cauvin, J. & Sanlaville, P.) 281–285 (CNRS, Paris 1981).
13. Stringer, C. B. *et al. Nature* **338**, 756–758 (1989).
14. Schwarcz, H. P. *et al. J. hum. Evol.* **17**, 733–737 (1988).
15. Grün, R., Stringer, C. B. & Schwarcz, H. P. *J. hum. Evol.* **20**, 231–248 (1991).
16. Grün, R. & Stringer, C. B. *Archaeometry* **33**, 2, 153–199 (1991).
17. Jelinek, A. J. in *The Emergence of Modern Humans* (ed. Mellors, P.) 81–90 (Edinburgh Univ. Press, Edinburgh, 1990).
18. Jelinek, A. J. in *The Evolution and Dispersal of Modern Humans in Asia* (ed. Akazawa, T.) (Univ. Mus. Bull. Univ. of Tokyo, in the press).
19. Rae, A. M. & Ivanovich, M. *Appl. Geochem.* **1**, 419–426 (1986).
20. McKinney, C. R. thesis, SMU Dallas (1991).
21. Grün, R. *et al. Nucl. Tracks* **14**, 237–241 (1988).
22. Garrod, D. & Bate, D. *The Stone Age of Mount Carmel* Vol. I (Oxford Univ. Press, Oxford, 1937).
23. Farrand, W. R. *J. archaeol. Sci.* **6**, 369–392 (1979).
24. McCown, T. D. in *The Stone Age of Mount Carmel* Vol. I (eds Garrod, D. & Bate, D.) 91–107 (Clarendon, Oxford, 1937).
25. Bar-Yosef, O. in *The Human Revolution: Behavioral and Biological Perspectives on the Origin of Modern Humans* (eds Mellor, P. & Stringer, C. B.) 589–610 (Edinburgh Univ. Press, 1989).
26. Oakley, K. P. *et al. Catalogue of Fossil Hominids III: Americas, Asia and Australia* (British Museum (Nat. History), London, 1975).
27. Stringer, C. B. & Trinkaus, E. in *Aspects of Human Evolution* (ed. Stringer, C. B.) 129–165 (Taylor and Francis, London, 1981).
28. Jelinek, A. J. in *Préhistoire du Levant* (eds Cauvin, J. & Sanlaville, P.) 265–280 (CNRS, Paris, 1981).
29. Masters, P. M. in *The Transition from Lower to Middle Palaeolithic and the Origin of Modern Man* (ed. Ronen, A.) 43–54 (Brit. Archaeol. Rep. Int. Ser. 151, Oxford, 1982).
30. Trinkaus, E. in *The Origin of Modern Humans* (eds Smith, F. H. & Spencer, F.) 251–293 (Liss, New York, 1984).
31. Edwards, R. L. *et al. Earth Planet. Sci. Lett.* **81**, 175–192 (1987).

ACKNOWLEDGEMENTS. We thank T. Bradshaw, M. Johnston, P. van Calsteren and D. Wright for assistance in the radiogenic isotope laboratory at the Open University.

Positive genetic correlation between female preference and preferred male ornament in sticklebacks

Theo C. M. Bakker

University of Bern, Zoologisches Institut, Abteilung Verhaltensökologie, Wohlenstrasse 50a, CH-3032 Hinterkappelen, Switzerland

A NUMBER of population genetics models predict the evolution of male sexual ornaments through female choice¹, but their genetic assumptions and predictions have hardly been investigated^{2,3}. A key feature of these models is a positive genetic correlation between male ornaments and female preference for them⁴. Here I test this prediction at the within-population level with three-spined sticklebacks, *Gasterosteus aculeatus*, which show conspicuous sexual dichromatism⁵. Intense red males are preferred in various situations^{6–10}, but there is great intrapopulation variation in redness both among wild-caught^{6,10} and among laboratory-bred males¹¹, which is partly environmental⁶ and may be partly genetic^{12,13}. Also, females show considerable intrapopulation variation in their preference for redder males^{6,8,9}, which is partly environmental^{8,9}. Wild-caught, intense red males and dull males were crossed with a number of females from the same population in a full-sib/half-sib breeding design. Daughters were tested for their preference for more intensely red males, and the sons' coloration was quantified. Both traits showed genetic variation. Also the redness of the sons correlated with the preference for red of their sisters, thus the two traits show positive genetic correlation.

Sexually mature sticklebacks were caught in the spring of 1990 from a Swiss freshwater population (near Roche/Montreux, 46°26' N, 6°55' E) which was introduced more than a century ago¹⁴. In the laboratory, fish were kept under simulated

summer conditions; males were housed singly in small tanks, whereas females were stored in female groups⁶. After nest building, males were used in sequential mate choice tests of ripe females⁸.

Six of the most extremely coloured males, that is, three intense red and three pale red males, served as fathers. They were crossed in a full-sib/half-sib breeding design¹⁵ with 14 females, which covered the whole spectrum of preference phenotypes for redder males as determined in sequential choice tests⁸. Paternal effects on offspring traits were excluded by removing clutches from the fathers' nests after fertilization. Progenies were raised and maintained in several small standardized full-sib groups per cross. Before the attainment of sexual maturity the sexes were separated; of each cross a random sample of males was housed individually and a random sample of females was maintained in standardized sister groups.

Two weeks after the completion of the first nest, the males' maximal intensity of red coloration on the throat was quantified¹⁶. Redder fathers produced on average significantly redder sons (Fig. 1), thus indicating additive genetic variation for red intensity. An analysis of variance (ANOVA) of red intensity of sons in full-sib and half-sib families also revealed a significant added variance component among fathers, indicating additive

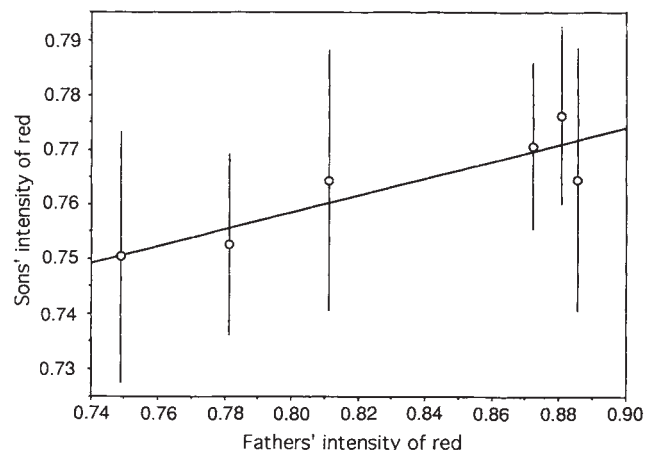


FIG. 1 Correlation between the intensity of red breeding coloration of wild-caught fathers and their laboratory-bred sons (average score of sons per father $\pm$ s.d.) ($r^2=0.79$, $F=15.02$, d.f. = 1, 4, $P=0.009$, 1-tailed). Number of tested sons (number of crosses) from left to right: 21(3), 7(1), 20(3), 9(2), 8(1), 38(5). Paternal effects were ruled out by removing clutches from the nests 1 h after fertilization, and hatching them artificially¹¹. The fish were raised in small standardized full-sib groups under simulated summer conditions (16:8 h in light/dark cycle, 15 °C) and fed freely. Because parasites might affect male coloration⁶, only food items were used that were most likely to be free of parasites. The few fish that caught or were suspected to have caught an *Oodinium* infection were not used in the tests. At first signs of developing breeding coloration, about 7 months after hatching, the sexes were separated, and a random selection of males individually housed⁶. Each row of 6 male tanks was illuminated by a 40 W fluorescent tube mounted 10 cm above the tanks. The males were regularly stimulated with ripe females, and most of them had built a nest within 2 months of isolation. Fathers' and sons' intensity of red was quantified 2 weeks after the completion of the first nest using Frischknecht's procedure¹⁶. Slides of the males were taken in a standardized set-up and analysed with a densitometer¹⁶. In the red throat region, the optical density of red (R, filter 700 nm), green (G, filter 546.1 nm), and blue (B, filter 435.8 nm) was measured at 10 defined points (diameter 0.5 mm). An appropriate measure of the intensity of red that is independent of the brightness of a colour, is the red index¹⁶, in which the R value (corrected for differences in film development) is expressed relative to the total colour density ($R+G+B$) and subtracted from 1 to obtain positive values between 0 and 1. The highest index for red on the throat was used in the analyses. A proof of the reliability of the method was obtained by a direct comparison¹⁰ of the red index with chroma calculated from reflectance spectra^{34,35} (Spearman rank correlation coefficient $r_s=0.80$, $N=15$, $P<0.0015$, 1-tailed).

TABLE 1 Sib analyses

(a) Sib analysis of variation in red coloration								
Source	d.f.	Full-sib and half-sib families			d.f.	Full-sib families		
		SS	F	P		SS	F	P
Between fathers	5	1,398,703	4.95	<0.02	5	808,872	2.03	<0.10
Between mothers (same fathers)	9	508,391	0.48	>0.75				
Within progenies	88	10,354,026			42	3,347,355		
(b) Sib analysis of variation in red preference								
Between fathers	5	2,049.13	4.48	<0.04	5	2,828.11	2.64	<0.05
Between mothers (same fathers)	8	731.88	0.52	>0.75				
Within progenies	47	8,348.70			31	6,641.62		

Sib analysis¹⁵ of variation in *a*, the intensity of red male coloration, and *b*, female preference for redder males. Red intensity data (red index $\times 100$) were square-transformed to meet the normality assumption for ANOVA (Kolmogorov-Smirnov test on transformed scores, $D=0.072$, $N=123$, P (Lilliefors)=0.112, 2-tailed)³³. The daughters' preferences were corrected for the degree to which their test males differed in red intensity by taking the residuals from the regression of preferences (*y*) on ranked differences in red intensity (*x*) between the test males of each pair ($y=56.54+2.88x$, $r^2=0.09$, $F=8.07$, d.f.=1, 84, $P<0.003$, 1-tailed). The distribution of standardized female preferences did not significantly deviate from normality (Kolmogorov-Smirnov test, $D=0.070$, $N=86$, P (Lilliefors)=0.346, 2-tailed)³³. The variation of red intensity of sons and of red preferences of daughters in full-sib and half-sib families were analysed with a two-level nested ANOVA with unequal sample sizes (left columns)³³. Variation among the most reliable full-sib family of each father (that is the one with the greatest number of scored sons or daughters) was analysed with a single classification ANOVA with unequal sample sizes (right columns)³³. Progeny was produced using a full-sib/half-sib breeding design¹⁵; over 2 weeks, 3 red and 3 dull males were crossed with 14 different females, 1–5 per male, which had been used the same day or the day before in sequential mate choice tests⁶. A few weeks after the first cross, three of the mothers became ripe again, and were crossed with a male from the other colour category. Because of the independence of the data set, their progeny was not used in the analyses except in the full-sib analysis and in the separate analyses for red and dull fathers (see below).

genetic variance of red intensity (Table 1a). Note that the variation among fathers is much greater than among their progenies (Fig. 1). Because the fathers were wild-caught, causes of variation in red intensity may be different for fathers and sons, making the regression unsuitable for heritability estimation. Narrow-sense heritability estimation from the sib analysis was not possible owing to the relatively low variance among mothers (Table 1a), which was probably effected by the use of extreme father coloration categories as opposed to the use of mothers from the whole spectrum of preference phenotypes. In addition, dominance effects may have hidden the mothers' contribution (see below). A full-sib analysis¹⁵ yielded a rough estimate of heritability of red intensity of 0.23 (s.e.¹⁷=0.27).

The preference of the female progeny for redder males was investigated with a simultaneous choice design⁶. When ripe, female progeny were given a choice between two courting males that differed in the intensity of red. Females were selected according to their ripeness and used once. When females were still ripe the day after the choice test, they were tested again to assess the repeatability¹⁵ of female choice.

Most females were consistent in their choice: they either preferred the redder male or showed no preference on both days

(Fig. 2). For unknown reasons a small number of females preferred the duller male on the first test day, but showed a clear preference for the redder male on the next (Fig. 2, empty circles). This group of females (relative preference for the redder male <35%: 11 out of 97 females tested, 6 daughters from red males and 5 from dull males) was left out in the following analyses, although their inclusion does not change the conclusions. Red preferences were significantly correlated between two successive days ($r^2=0.41$, $F=11.34$, d.f.=1, 16, $P=0.004$). The repeatability (coefficient of intraclass correlation) of preference was estimated as 0.65 ± 0.14 (s.e.¹⁷) (single classification ANOVA, $F=4.73$, d.f.=17, 18, $P<0.001$).

An ANOVA of female preference within and between full-sib and half-sib families indicated the presence of additive genetic variance of preference for redder males, because the added variance component among fathers was significant (Table 1b). The relatively low variance of the daughters' preference among mothers (Table 1b) prevented narrow-sense heritability estimation. A full-sib analysis gave a rough estimate of heritability of preference for redder males of 0.43 (s.e.¹⁷=0.37).

The influence of mothers both on their daughters' red preference and on their sons' red intensity tended to be stronger in

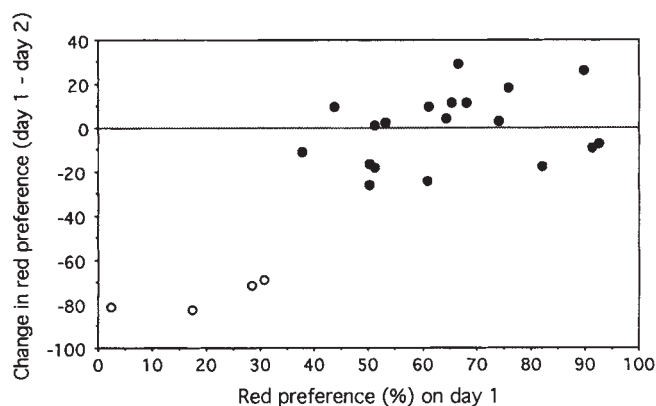


FIG. 2 Change in preference for the redder male of 22 ripe females between two successive days (day 1–day 2) as a function of the preference on day 1. Preference is expressed as % of total duration of head-up display directed at the redder male. The horizontal line indicates no change. Females with a preference less than 35% on day 1 are indicated by empty circles. Females were maintained together with their brothers in several small standardized groups per cross. Before the attainment of sexual maturity they were separated from their brothers by placing 15 randomly chosen sisters of each cross in 60-l tanks. Each three female tanks was lit by a 40 W fluorescent tube lying on top of the tanks. Ripe females were selected out of the sister groups, each was isolated in a small tank, and tested the next day for their preference for redder males. In the choice test, females were enclosed in a plexiglass cell and offered a simultaneous choice between two courting males that differed in their intensity of red coloration, and could not interact with each other⁶. The duration of the female's head-up courtship posture while pointing at one of the two males was measured during a 2 min period after 1 min of acclimatization. This measure correlates positively with the probability to spawn with that male⁷. Three different dull and four different red males were used in five different combinations. Positions of the males were regularly changed. After the choice test, females were put back in their individual tanks and after spontaneous spawning they were marked by clipping the tip of one or more spines and put back in their sister groups.

crosses with dull fathers than when crossed with red males (between mothers source of variation of daughters' preference in crosses with red males: $F=0.50$, d.f. = 5, 26, $P>0.77$; and in crosses with dull males: $F=1.81$, d.f. = 6, 43, $P<0.12$; for sons' coloration, $F=0.61$, d.f. = 5, 47, $P>0.69$, and $F=1.34$, d.f. = 7, 58, $P<0.25$, respectively). These results may indicate (partial) dominance of genes that promote preference for redder males and intense red coloration.

A positive genetic correlation between ornament and preference was indicated by the significant positive correlation between the fathers' intensity of red and their daughters' average preference for red (preferences corrected for the degree to which test males differed in their red intensities; $r^2=0.77$, $F=13.30$, d.f. = 1, 4, $P=0.011$, 1-tailed). Conclusive evidence of this positive genetic correlation is given by the highly significant positive correlation that existed across fathers between the sons' intensity of red and the daughters' preference for redder males (Fig. 3). The phenotypic correlation of 0.998 roughly estimates the genetic correlation between male and female traits because by using group means environmental effects are reduced. A more precise estimate of 0.75 (s.e. $^{15,18}=0.31$) was calculated from the slope of the regression line (using red index $\times 100$ and average scores per cross averaged over the different crosses per father: $y=-651.13+8.507x$, $r^2=0.998$, $F=1709.01$, d.f. = 1, 4, $P<0.0001$) and the estimated heritabilities of the male and female trait 15,19 .

It is unlikely that the strong correlation between male and female traits (Fig. 3) is caused by nonheritable effects 20 : common environmental influences were minimized by the applied methods of raising and maintaining fish and were apparently low as indicated by the relatively small variance of both traits between mothers crossed with the same father (Table 1). Precise estimates of genetic variances and covariances require large sample sizes 15 and may thus pose practical problems in quantitative genetic studies of sexual selection. This study clearly demon-

strates significant genetic variation in male ornament and female preference and positive genetic correlation between them, but precise estimates of the genetic variables cannot be obtained with such data sets: the estimates may be inflated by the use of full-sib comparisons, the estimated genetic correlation has a large standard error, and the heritability estimates are not significant.

Data on genetic variation in female mating preference and genetic covariation between female preference and preferred male traits are scarce 2,3 and partly ambiguous $^{21-23}$. These genetic variables are critical in population genetics models of sexual selection. During the evolution of male ornaments, both 'fisherian' and 'good genes' models predict a positive genetic correlation between preference and male ornament. At equilibrium, in the Fisher process a positive genetic covariance will typically be maintained in both major gene 24,25 and polygenic models $^{19,26-28}$. The genetic covariance can even be maintained when female choice is costly 29 , which is likely in this stickleback population 9 . In major gene models of 'good genes' sexual selection, there are no internal equilibria, and the male and/or female trait will become fixed 1 . But in polygenic models it is usually assumed that genetic variation in both traits is maintained and that there is a positive genetic covariance at equilibrium.

Thus one cannot distinguish whether a fisherian or good genes process, alone $^{19,24-30}$ or in combination 31,32 , maintains genetic variation and covariation of the evolved male and female traits in this stickleback population. My findings confirm the genetic predictions of models of sexual selection, but cannot exclude alternative hypotheses without further investigations. For example, the possibility of multiple introductions of sticklebacks in these waters, either through repeated introductions from one source population or a single introduction from more than one source, is difficult, if not impossible, to rule out, and may have caused a transitory genetic correlation between male and female traits. □

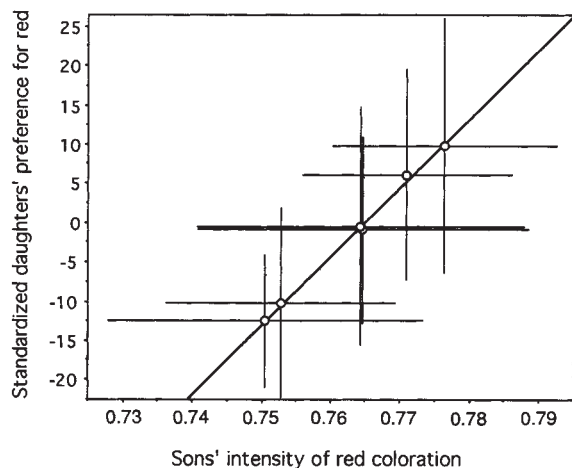


FIG. 3 Correlation across fathers ($N=6$) between the average ($\pm$ s.d.) red intensity of sons ($N=103$ sons, for numbers of sons and crosses per father, see Fig. 1) and the average ($\pm$ s.d.) preference of daughters for redder males ($N=61$ daughters, numbers of tested daughters per father from left to right: 7, 5, 15, 25, 5, 4). The daughters' preferences were corrected for the degree to which their test males differed in their red intensities. The regression equation is: $y=-667.60+872.94x$ ($r^2=0.997$, $F=1038.20$, d.f. = 1, 4, $P<0.0001$). In producing offspring, matings were forced, that is I determined beforehand to which male a particular ripe female would be released. Criteria that were used in making up mating pairs were that both categories (red and dull) of males could reproduce about equally often, and that within each category the males were crossed with the whole spectrum of relative female preferences for intense red males, as had been determined in the sequential choice tests 8 . With a few exceptions, the females spawned with their imposed males.

Received 11 December 1992; accepted 26 February 1993.

- Maynard Smith, J. *Trends Ecol. Evol.* **6**, 146–151 (1991).
- Heisler, L. et al. in *Sexual Selection: Testing the Alternatives* (eds Bradbury, J. W. & Andersson, M. B.) (Wiley, Chichester, 1987).
- Bakker, T. C. M. *Neth. J. Zool.* **40**, 617–642 (1990).
- Kirkpatrick, M. & Ryan, M. J. *Nature* **350**, 33–38 (1991).
- Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
- Millinski, M. & Bakker, T. C. M. *Nature* **344**, 330–333 (1990).
- McLennan, D. A. & McPhail, J. D. *Can. J. Zool.* **68**, 482–492 (1990).
- Bakker, T. C. M. & Millinski, M. *Behav. Ecol. Sociobiol.* **29**, 205–210 (1991).
- Millinski, M. & Bakker, T. C. M. *Proc. R. Soc. B* **250**, 229–233 (1992).
- Bakker, T. C. M. & Mundwiler, B. *Behav. Ecol.* (in the press).
- Bakker, T. C. M. *Behaviour* **98**, 1–144 (1986).
- Hagen, D. W. & Moodie, G. E. E. *Evolution* **33**, 641–648 (1979).
- Bakker, T. C. M. in *The Evolutionary Biology of the Threespine Stickleback* (eds Bell, M. A. & Foster, S. A.) (Oxford Univ. Press, Oxford, in the press).
- Laurent, P. J. *J. Fish. Res. Bd. Canada* **29**, 867–875 (1972).
- Falconer, D. S. *Introduction to Quantitative Genetics* 3rd edn (Longman, Harlow, 1989).
- Frischnecht, M. *Evol. Ecol.* (in the press).
- Becker, W. A. *Manual of Quantitative Genetics* 4th edn (Academic Enterprises, Pullman, 1985).
- Robertson, A. *Biometrics* **15**, 469–485 (1959).
- Landé, M. *Proc. natn. Acad. Sci.* **78**, 3721–3725 (1981).
- van Noordwijk, A. J. in *Population Biology and Evolution* (eds Wöhrman, K. & Loeschke, V.) (Springer, Heidelberg, 1984).
- Kearns, P. W. E., Tomlinson, I. P. M., Veltman, C. J. & O'Donald, P. *Heredity* **68**, 385–389 (1992).
- O'Donald, P. J. & Majerus, M. E. N. *Heredity* **69**, 521–526 (1992).
- Ritchie, M. G. *Trends Ecol. Evol.* **7**, 328–329 (1992).
- Kirkpatrick, M. *Evolution* **36**, 1–12 (1982).
- Seger, J. *Evolution* **39**, 1185–1193 (1985).
- Heisler, I. L. *Heredity* **55**, 187–198 (1985).
- Kirkpatrick, M., Price, T. & Arnold, S. J. *Evolution* **44**, 180–193 (1990).
- Barton, N. H. & Turelli, M. *Genetics* **127**, 229–255 (1991).
- Pomiankowski, A., Iwasa, Y. & Nee, S. *Evolution* **45**, 1422–1430 (1991).
- Iwasa, Y., Pomiankowski, A. & Nee, S. *Evolution* **45**, 1431–1442 (1991).
- Kirkpatrick, M. *J. theor. Biol.* **119**, 263–271 (1986).
- Tomlinson, I. P. M. *Heredity* **60**, 283–293 (1988).
- Sokal, R. R. & Rohlf, F. J. *Biometry* 2nd edn (Freeman, New York, 1981).
- Endler, J. A. *Biol. J. Linn. Soc.* **41**, 315–352 (1990).
- Endler, J. A. *Vision Res.* **31**, 587–608 (1991).

ACKNOWLEDGEMENTS. I thank M. Frischnecht and B. Mundwiler for help in quantifying male coloration, M. Millinski and A. Pomiankowski for helpful comments, W. Rowland (Indiana University) for facilities, C. Baube for allowing me to measure and analyse reflectance spectra, A. Pomiankowski, A. van Noordwijk, C. Wedekind and F. Weissing for discussions, A. van Noordwijk for statistical advice, and the Swiss National Science Foundation and the Swiss Academy of Science for financial support.

African-European honeybee hybrids have low nonintermediate metabolic capacities

Jon F. Harrison* & H. Glenn Hall†

* Department of Zoology, Arizona State University, Tempe, Arizona 85287-1501, USA

† Department of Entomology and Nematology, University of Florida, Gainesville, Florida 32611-0620, USA

BECAUSE of a number of behavioural, ecological and physiological factors, African honeybees (*Apis mellifera scutellata*) are better adapted to tropical environments than European bees. African bees achieve higher rates of reproduction and colony growth partly by collecting more pollen relative to nectar¹ and by allocating more nutrients to brood rearing². African colonies generate more swarms^{3,4} which travel far greater distances⁵. These adaptive traits have been dramatically manifested in the neotropics: African bees have formed large expanding feral populations, whereas few European bees have survived outside apiaries⁶. The feral populations consist of African matrilines⁷⁻⁹ that have not hybridized extensively with European drones from apiaries¹⁰. Perhaps because of their European heritage, hybrids are poorly adapted to the tropics and thus do not survive. European matrilines, however, do not become part of the feral population, even after repeated backcrossing with feral African drones⁷⁻¹⁰. Conceivably, negative heterosis could further reduce hybrid survival, where European maternal factors are particularly detrimental⁷. We report here physiological findings consistent with these possibilities. Mass-specific metabolic capacities were higher in African bees than in European bees and were low and nonintermediate in hybrids. Thus, higher metabolic and flight capacities may contribute to African bees' competitive advantages in the tropics.

The faster developmental period², smaller body size^{2,11}, shorter lifespan^{2,12}, similar thorax temperature during flight¹³ and greater swarming distances⁵ of African bees suggest that they have greater mass-specific metabolic rates (MR) than European bees. Nonflying, cold-exposed groups of African bees do indeed have higher specific MR¹⁴. During flight, a higher specific metabolic capacity, defined as the maximal MR per gram of body, could enhance many aspects of flight and foraging performance, including manoeuvrability, lifting capacity and maximal flight speeds¹⁵. To determine if there were differences in metabolic capacity, we compared CO₂ production during agitated flight among parental African and European honeybee stock, and reciprocal, first generation and backcross hybrids (groups listed in Fig. 1).

After the MR measurements, bees were frozen and body and thorax mass determined. The weights agreed with the well known smaller size of African compared to European honeybees (Fig. 1). The hives used for this study contained mostly combs with the larger European size cells. Thus, the average size of the African bees was somewhat larger than if they had been reared in only African combs. Hybrid bees were intermediate to European-like in mass (Fig. 1).

The MR of European bees found in this study were equal to the highest ever measured for honeybees and were similar to those reported for maximally loaded workers¹⁶, supporting the conclusion that these protocols measured metabolic capacity (Fig. 2). African bees had greater mass-specific MR than European bees (Fig. 2). The greater thorax-specific metabolic capacity in African bees suggests a higher oxidative capacity in their flight muscle relative to European and hybrid bees. Significantly, the MR of hybrid bees were not intermediate but, in all instances, were equal to or lower than those measured for European bees (Fig. 2).

Smaller animals use more energy to hover¹⁵; however, the

differences in MR of African, European and hybrid bees were not explicable by differences in body size alone. African bees had higher, and hybrid bees had lower, MR than predicted by their body masses (Fig. 3). These data suggest that African bees are superior, and hybrid bees are inferior, to European bees in maximal flight performance. Differences in maximal flight performance would affect foraging, hive defence and dispersal and thus could contribute to differences in fitness among the genotypes.

Possibly the differences in MR observed in this study reflect differences in temperament among the genotypes, with African bees responding more strongly, and hybrid bees less strongly than European bees to the agitation stimulus. Because loading does not elicit significantly higher MR¹⁶, it seems likely that the response of the European workers is at or near its maximal level. Furthermore, African-European hybrid bees are intermediate to African-like in defensive characteristics¹⁷⁻¹⁹, indicating that depressed irritability is not the mechanism responsible for depressed metabolic capacity in these hybrid bees.

Our results also suggest that the inheritance of MR in African-European honeybee hybrids is nonreciprocal. MR for hybrid progeny of European queens × African drones were significantly lower than progeny of African queens × European drones (a

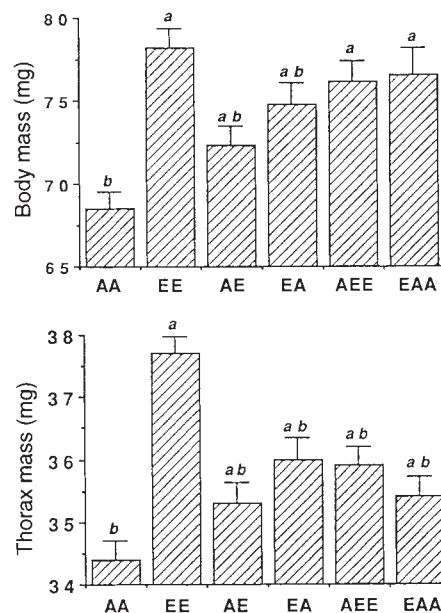


FIG. 1 Body mass (top) and thorax mass (bottom, including wings and legs) of worker progeny from the following crosses: (AA) African queens × African drones, (EE) European queens × European drones, (AE) African queens × European drones, (EA) European queens × African drones, (AEE) queen daughters from AE crosses × European drones, (EAA) queen daughters of EA crosses × African drones. Means and s.e.s shown. 'a' Significant difference from AA bees; 'b', the group differed from EE bees (*a priori* contrasts, overall $P < 0.05$ calculated using the Bonferroni procedure²⁰).

METHODS. This study was conducted in March, 1992 at the Escuela Agrícola Panamericana in Zamorano, Honduras. Parental stocks were *A.m. scutellata* (feral Honduran swarms) and *A.m. ligustica* (Kona Queen Company, Hawaii). Coming from New World populations, neither of the stocks can be considered as pure subspecies. Several colonies served as different sources of queens and drones, but five to six drones from the same colony were used to inseminate each queen. Inbreeding was avoided in the backcrosses. Colonies were randomly oriented and widely spaced to reduce drifting of foragers. Each bee was analysed for its mitochondrial DNA type⁹, which enabled elimination of drifters of the opposite matriline, but not between parental and hybrid colonies of the same matriline. We assumed that errors caused by drifters would have reduced the differences among the groups. Final sample sizes consisted of 7 to 21 workers from each of 4 colonies per group, except that the EAA group had only 3 colonies. A final total of 317 bees were used for the MR determinations.

priori contrast²⁰, pooled EA+EAA versus AE+AEE, $P=0.024$).

Hybrid honeybees are formed as the invading African population expands into regions occupied by European colonies, but the hybrids do not persist as the African population becomes established¹⁰. In hybrids, some defensive and morphological characteristics are intermediate in expression^{17,18,21}. If hybrid fitness were intermediate in the tropical environment, this could account for the apparent loss of hybrid bees. Hybrids would still be expected to survive in managed apiaries, as have poorly adapted European bees^{10,21,22}, and in subtropical-temperate transition zones where African and European bees may be equally adapted²²⁻²⁴. Hybrid presence in some environments does not exclude the possibility that nonadditive genetic factors could also reduce hybrid fitness. If adaptive factors were strictly additive, swarms generated from European matrilines, repeatedly backcrossed to African drones, should have become sufficiently adapted to survive in the neotropical environment. However, a European maternal factor seems to remain as a selective disadvantage. (Genetic processes involved in African bee spread have been reviewed recently^{25,26}.)

Our data suggest that negative heterosis for MR may mediate some aspect of depressed hybrid fitness of feral colonies in the neotropics. A possible cause of negative heterosis is a disruption of coadapted enzymatic complexes in hybrids^{7,25,26}. The low

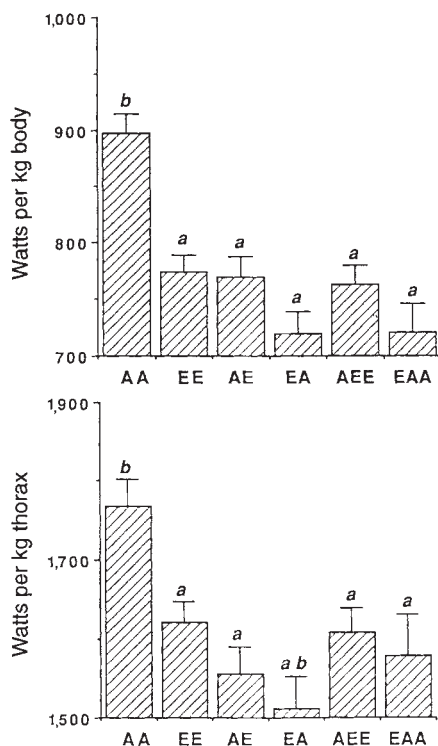


FIG. 2 Mass-specific metabolic capacities, per gram body (top) and per gram thorax (bottom) of workers from African, European and hybrid honeybee colonies. Symbols the same as in Fig. 1.

METHODS. The mass-specific metabolic capacity (watts kg^{-1}) was calculated by measuring the carbon dioxide production of individual bees in agitated flight, using a flow-through respirometry system²⁷. The speed of response was optimized using a LICOR LI-5152 CO_2 analyser and a high flow rate (1.9 l min^{-1} , regulated to $\pm 1 \text{ ml min}^{-1}$ by a mass-flow meter) relative to chamber size (200 ml). Out-going foragers were placed into a plexiglass chamber within 2 min of capture. The chamber walls were hit to induce the bees to fly and CO_2 production measured over a 1 min period. Similar MRs were measured over 15–30 s periods, but MR (and flight behaviour) tended to decline if bees were kept in the containers for several minutes, or if the bees were undisturbed. Carbon dioxide production rates were converted to MR assuming a respiratory quotient equal to one²⁸.

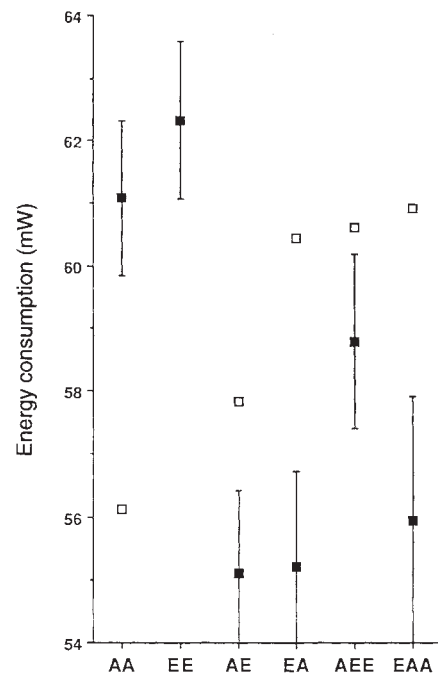


Fig. 3 Energy consumption (mW, mean and s.e., filled symbols) and predicted energy consumption (open symbols) of the honeybee groups described in Fig. 1. Predicted values were calculated using European bees as the standard, assuming the MR of the African and hybrid groups should differ from European bees in relation to body mass^{0.62}, as shown for euglossine bees²⁹, and European honeybee workers and drones³⁰.

MR in African-European honeybee hybrids from reciprocal crosses indicates that a European maternal component is not completely responsible. Because MR are lower in hybrids from European matrilines, disrupted complexes might include those with nuclear and mitochondrial-encoded subunits, where European mitochondrial-encoded units have a greater negative effect in heterologous combinations.

Our findings suggest that differences in flight capacity or metabolism may be an important component of fitness differences among honeybee subspecies. Furthermore, the nonadditive and nonreciprocal inheritance of flight metabolism in African-European hybrids may be a selective disadvantage that has enhanced preservation of the African genotype during colonization of tropical America. □

Received 21 December 1992; accepted 25 January 1993.

1. Danka, R. G., Hellmich, R. L. II, Rinderer, T. E. & Collins, A. M. *Anim. Behav.* **35**, 1858–1863 (1987).
2. Winston, M. L. in *The 'African' Honey Bee* (eds Spivak, M., Fletcher, D. J. C. & Breed, M. D.) 201–212 (Westview, Boulder, 1991).
3. Winston, M. L., Drotkyn, J. & Taylor, O. R. *Oecologia* **48**, 407–413 (1981).
4. Otis, G. W. in *Social Insects of the Tropics* Vol. I (ed. Jaisson, P.) 209–219 (Université Paris-Nord, Paris, 1982).
5. Fletcher, D. J. A. *Rev. Ent.* **23**, 151–171 (1978).
6. Michener, C. D. A. *Rev. Entomol.* **20**, 399–416 (1975).
7. Hall, H. G. & Muralidharan, K. *Nature* **339**, 211–213 (1989).
8. Smith, D. R., Taylor, O. R. & Brown, W. W. *Nature* **339**, 213–215 (1989).
9. Hall, H. G. & Smith, D. R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4548–4552 (1991).
10. Hall, H. G. *Genetics* **125**, 611–621 (1990).
11. Winston, M. L., Taylor, O. R. & Otis, G. W. *Bee World* **64**, 12–21 (1983).
12. Winston, M. L. & Katz, S. J. *Can. J. Zool.* **59**, 1571–1575 (1981).
13. Heinrich, B. J. *exp. Biol.* **80**, 217–229 (1981).
14. Southwick, E. E., Roubik, D. W. & Williams, J. M. *Comp. Biochem. Physiol.* **A97**, 1–8 (1990).
15. Casey, T. M. in *Insect Flight* (eds Goldworthy, G. J. & Wheeler, C. H.) 257–272 (CRC Press, Boca Raton, FL, 1989).
16. Wolf, T., Schmid-Hempel, P., Ellington, C. P. & Stevenson, R. D. *Funct. Ecol.* **3**, 417–424 (1989).
17. Collins, A. M. & Rinderer, T. E. in *The 'African' Honey Bee* (eds Spivak, M., Fletcher, D. J. C. & Breed, M. D.) 309–328 (Westview, Boulder, 1991).
18. Stort, A. C. & Gonçalves, L. S. in *The 'African' Honey Bee* (ed. Spivak, M., Fletcher, D. J. C. & Breed, M. D.) 329–358 (Westview, Boulder, 1991).
19. Guzmán-Novoa, E. & Page, R. E. *Ann. Ent. Soc. Am.* (in the press).
20. Wilkinson, L. *SYSTAT: the System for Statistics* (SYSTAT Inc., Evanston, IL, 1990).

21. Rinderer, T. E., Stelzer, J. A., Oldroyd, B. P., Buco, S. M. & Rubink, W. L. *Science* **253**, 309–311 (1991).
22. Taylor, O. R. *Bull. Ent. Soc. Am.* **31**, 14–24 (1985).
23. Sheppard, W. S., Rinderer, T. E., Mazzoli, J. A., Stelzer, J. A. & Shimanuki, H. *Nature* **349**, 782–784 (1991).
24. Lobo, J. A., Lama, M. A. & Mestriner, M. A. *Evolution* **43**, 794–802 (1989).
25. Hall, H. G. *Florida Ent.* **75**, 51–59 (1992).
26. Hall, H. G. in *The 'African' Honey Bee* (eds Spivak, M., Fletcher, D. J. C. & Breed, M. D.) 45–76 (Westview, Boulder, 1991).
27. Harrison, J. F., Phillips, J. E. & Gleeson, T. T. *Physiol. Zool.* **64**, 451–472 (1991).
28. Beenakkers, A. M. Th., Van Der Horst, D. J. & Van Marrewijk, W. J. A. *Insect Biochem.* **14**, 243–260 (1984).
29. Casey, T., May, M. L. & Morgan, K. R. *J. exp. Biol.* **116**, 271–289 (1985).
30. Coelho, J. R. *Environ. Ent.* **20**, 1627–1635 (1991).

ACKNOWLEDGEMENTS. We thank the members of the Escuela Agrícola Panamericana, Zamorano, Honduras, for their help with this project, especially R. Villavicencio, A. Suazo and J. Quan for establishing and maintaining the honeybee colonies and assisting with the collections, the administrative staff, R. Salas (Apiculture), A. Montes (Chairman-Horticulture) and S. Malo (Director), for support and J. Fewell, S. Rissing, W. Doane and K. Helms for comments on the manuscript. This work was supported by grants from the US National Science Foundation and a FGA award from Arizona State University to J.F.H., and a USDA NRI Competitive Grant and US-AID Program Support Grant to H.G.H.

Antisense oligodeoxynucleotides to NMDA-R1 receptor channel protect cortical neurons from excitotoxicity and reduce focal ischaemic infarctions

C. Wahlestedt*, E. Golanov*, S. Yamamoto*, F. Yee*, H. Ericson*, H. Yoo*, C. E. Inturrisi† & D. J. Reis*

* Division of Neurobiology, Department of Neurology and Neuroscience, and † Department of Pharmacology, Cornell University Medical College, 411 East 69th Street, New York, New York 10021, USA

THE excitatory amino acid, L-glutamate, acting through its N-methyl-D-aspartate (NMDA) receptor, may contribute to neuronal death following cerebral vascular occlusion^{1–3}. In support of this hypothesis, NMDA receptor antagonists reduce the volume of infarction produced by occlusion of the middle cerebral artery *in vivo*^{4,5} and attenuate Ca²⁺ influx and neuronal death elicited by L-glutamate or NMDA *in vitro*^{3,6}. A complementary DNA coding for a major component of the NMDA receptor channel complex, a single protein of *M_r* 105.5K (NMDA-R1), has been isolated from rat brain⁷. Here we demonstrate that inhibition of the synthesis of NMDA-R1 by treatment with antisense oligodeoxynucleotides selectively reduces the expression of NMDA receptors, prevents the neurotoxicity elicited by NMDA *in vitro* and reduces the volume of the focal ischaemic infarction produced by occlusion of the middle cerebral artery in the rat.

Several phosphodiester antisense oligodeoxynucleotides (ODNs) corresponding to untranslated and translated amino-terminal regions of NMDA-R1 messenger RNA^{7–9} were synthesized. An 18-mer, NMDAR1AS/c, designed to correspond to nucleotides 4–21 (ref. 7), which directly follow the translation initiation codon, was most effective *in vitro* and was subsequently used for intracerebroventricular (i.c.v.) administration in the rat. The *in vitro* experiments (Fig. 1) used a combination of NMDAR1AS/c and two other ODNs (described in Fig. 1 legend). Control cultures and rats were treated with matching sense or missense (that is, scrambled antisense) 18-mers or vehicle.

First we investigated whether treatment of neuron-enriched cultures with an antisense ODN to the NMDA-R1 would: (1) inhibit binding of a selective ligand for the NMDA receptor, [¹²⁵I]iodo-dizocilpine (MK801); (2) block NMDA-mediated influx of Ca²⁺; and (3) protect neurons from destruction by exposure to the excitotoxin, NMDA. Neurons were isolated from the cerebral cortex of rat embryos on embryonic day 14

(E14) and then cultured for up to 21 days under conditions that curtail proliferation of non-neuronal cells¹⁰. To promote stability of ODNs, serum was omitted from the tissue culture medium after 4 days *in vitro*. ODNs were added at that time and every 2 days thereafter (20%, v/v). Treatment consisted of 1 μM of each of three NMDA-R1 antisense, sense or missense ODNs (Fig. 1). An untreated culture served as control.

The density (*B_{max}*) of NMDA receptor binding sites, as determined in whole culture by the NMDA-selective ligand [¹²⁵I]iodo-MK801 (ref. 11), was reduced to 56% of the *B_{max}* (0.74 ± 0.06 pmol per mg protein; mean ± s.e.m.) of untreated cultures (compare Fig. 1a) by treatment with the antisense ODNs. The affinity of the ligand was unaffected (result not shown). Treatment with sense or missense ODNs had no effect. The effect of NMDA-R1 antisense ODNs seems to be specific for the NMDA receptor because treatment did not modify the binding sites of either of two ligands that are unrelated to the NMDA receptor channel complex: (1) [¹²⁵I]peptide YY (using intact neurons, as in ref. 12; *n* = 6; data not shown) and (2) [³H]-6-cyano-7-nitroquinoxaline-2,3-dione, CNQX (binding was assayed on washed membranes from cultured cells at the reported *K_d* of 39 nM (ref. 13); *n* = 4; data not shown).

Next, we investigated whether calcium ion flux through the NMDA-gated channel was altered. Treatment with antisense ODNs inhibited the NMDA receptor-mediated influx of ⁴⁵Ca²⁺ into 14-day-old cultures by 48% (Fig. 1b).

Treatment with antisense ODNs for the NMDA-R1 also protected against NMDA-induced neurotoxicity. Thus, 18–20 h after exposure of 20-day-old neuronal cultures to NMDA (300 μM) for 30 min, 60–70% of treated neurons in control cultures were destroyed and did not exclude a marker dye. Pretreatment with NMDA-R1 antisense ODNs preserved more than 50% of the NMDA-sensitive neurons (Fig. 1c), whereas treatment of sister cultures with sense or missense ODN sets were not protected from the excitotoxin. (The neuronal viability, also monitored by lactate dehydrogenase release throughout the experiments, was unaffected at total ODN concentrations of 3 μM. But when the cultures were exposed to concentrations 10-fold higher (30 μM) of any of the three sets of ODNs (antisense, sense, or missense) the number of viable neurons was slowly reduced over several days, possibly indicating a non-specific toxic action of the reagents. Moreover, a phosphorothioate analogue of NMDAR1AS/c was not tolerated well by the neurons at 30 μM.

On the basis of the *in vitro* experiments, we investigated whether an antisense ODN for the NMDA receptor could reduce the size of focal ischaemic infarctions induced by occlusion of the middle cerebral artery *in vivo*. Groups of rats of the spontaneously hypertensive strain SHR (so selected because of the uniformity of lesions within and between groups; see ref. 14) were anaesthetized with halothane (1.8–2.5%), and indwelling cannulas were inserted into the lateral ventricle aseptically and the wounds closed. After recovery, an antisense ODN, NMDAR1AS/c, or the corresponding sense ODN (15 nmol for each in 5 μl), or vehicle was administered i.c.v. twice daily for two days. Forty-eight hours after the first injection, animals were reanaesthetized with halothane, the middle cerebral artery was occluded, the wounds were closed and animals returned to their cages. Antisense-ODN-treated animals did not display any behavioural or neurological abnormalities upon gross examination. Twenty-four hours later animals were killed, their brains removed and sectioned, and the distribution and volume of the infarctions measured using an image analysis system¹⁴. A fourth group of rats received MK801 (1 mg kg⁻¹ intravenously) 30 min after occlusion of the middle cerebral artery. Systemic blood pressure was reduced by MK801 but not by the ODNs (see legend to Fig. 2).

Treatment with either the antisense ODN or MK801 significantly reduced lesion volume, the former by 43.5% and the latter by 31.4% over controls (Fig. 2). As shown in Fig. 3,

distribution of salvage with either treatment was predominantly confined to the posterior area of infarction and included portions of occipital cortex and lateral caudate nucleus; this distribution corresponds to regions salvaged by other treatments and to the ischaemic penumbra¹⁴⁻¹⁶. The sense ODN and vehicle were without effect (Figs 2 and 3).

To determine whether antisense ODN treatment altered concentrations of cortical NMDA receptors and/or NMDA-R1 mRNA, separate groups of SHR rats were treated identically, receiving antisense (NMDAR1AS/c) or sense ODN, or saline. About 54 h after the first injection, the rats were killed. The frontal pole of the cerebral cortex was removed and processed for measurement of NMDA-R1 mRNA by solution hybridization¹⁷ and the remaining cortical tissue was used for assessment of receptor density with the high-affinity competitive NMDA receptor antagonist, [³H]CGS 19755 (ref. 18).

Treatment with the antisense ODN significantly reduced (by

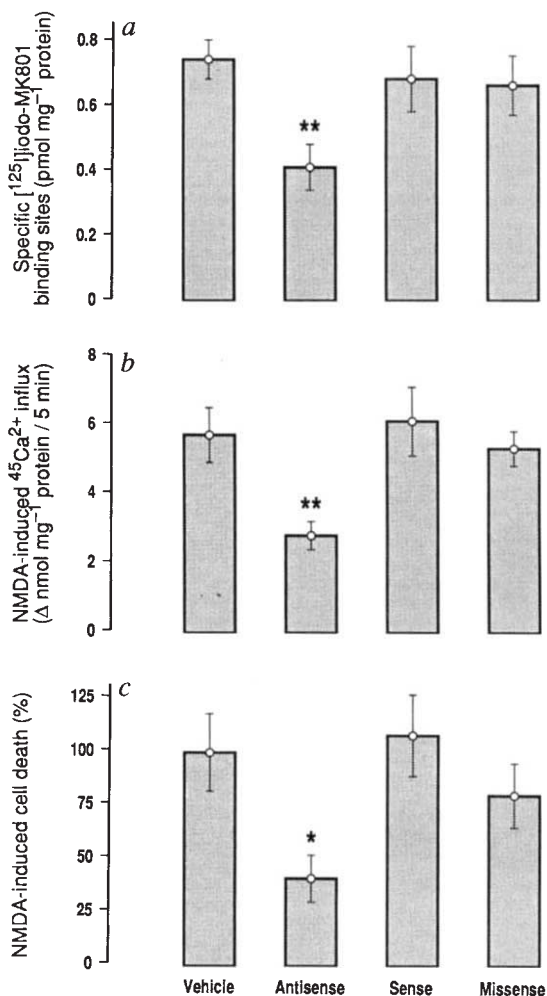
30–40% compared with sense or vehicle) the density ($B_{\max}$) of cortical binding sites for [³H]CGS 19755 (Fig. 4) while not affecting the apparent binding affinity (results not shown). In contrast, the steady-state concentration of cortical NMDA-R1 mRNA was unchanged by antisense ODN treatment (Fig. 4). Thus antisense ODN seems to act primarily to suppress NMDA-R1 protein translation selectively. Our data do not support major participation by RNase H (ref. 19).

To exclude the possibility that unrelated receptor proteins were downregulated by the antisense ODN, we studied binding of two other radioligands, [¹²⁵I]-labelled peptide YY (ref. 12) and [³H]CNQX (ref. 13), with cerebral cortex and cerebellum. Treatment with antisense ODN did not modify the binding site density of [¹²⁵I]peptide YY nor [³H]CNQX (results not shown).

Finally, as an additional approach for *in vivo* treatment of the rat, a phosphorothioate analogue of NMDAR1AS/c was administered continuously (80 nmol over 72 h by osmotic

FIG. 1 Downregulation of functional NMDA-R1 receptor channels and neuroprotection by antisense ODN treatment, to NMDA-R1, of rat cortex neuron-enriched cultures *in vitro*. Cultures were studied with respect to *a*, specific [¹²⁵I]iodo-MK801 binding sites; *b*, NMDA-induced ⁴⁵Ca²⁺ influx; and *c*, NMDA-induced neuronal death. Parallel cultures received no (control), sense, or missense (scrambled antisense) ODNs.

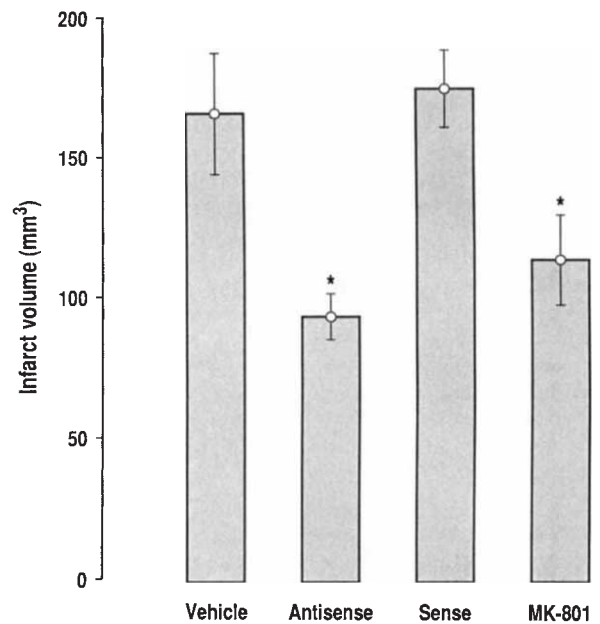
METHODS. Cultures were prepared from E14 rat embryo cerebral cortex. Following dissociation in 0.025% trypsin (Worthington), cells were plated on polyornithine-coated (Sigma) Thermanox coverslips (Nunc) placed in 20-mm Nunc (Nunc) plastic tissue culture wells, or directly into polyornithine-coated 35-mm wells. Culture medium was Dulbecco's modified Eagle's medium (DMEM; Sigma) with 10% bovine calf serum (Irvine), 2 mM L-glutamine, 10.5 mg ml⁻¹ glucose (final concentration), 100 µg ml⁻¹ streptomycin and 100 U ml⁻¹ penicillin (Sigma). After 4 days, serum was invariably omitted and ODNs were added; coverslips were inverted (a procedure that improved long-term neuronal survival and toxicity analyses) and placed in wells containing DMEM with N2 supplements²⁰. Cultures received a set of three antisense ODNs, that is, NMDAR1AS/a (5'-TTCATGCTGCGCTGCT-3', corresponding to nucleotides 100–117 of ref. 7), NMDAR1AS/b (5'-CATGAG-CTCCGGGCACAG-3', -15–+3), and NMDAR1AS/c (5'-CAGCAGGTGCAT-GGTGCT-3', 4–21). Control cultures received no ODNs while yet other cultures received sets of either three sense ODNs or three missense ODNs, based on the antisense sequences. The ODNs were then present throughout the culture periods. The medium was not subsequently replaced because regular medium changes caused reduced neuronal survival. Instead, every 2 days the medium of the wells was fortified by 20% (v/v) of fresh medium with or without ODNs. It is known that ODNs are quite stable in serum-free conditions^{12,19}. Cultures were subsequently subjected to three types of experiment: *a*, Radioreceptor binding studies were performed on 12-day-old intact cells, using [¹²⁵I]iodo-MK801 (New England Nuclear). Radioligand binding assays were based on previously described methods^{11,12}, described here briefly. For intact cell binding assays, cultured cells were washed 3 times with binding buffer (141 mM NaCl, 20 mM HEPES, 5.4 mM KCl, 0.44 mM KH₂PO₄, 1.26 mM CaCl₂, 0.81 mM MgSO₄, 10 mM glucose; pH 7.4; buffer A). Incubations of the intact cells with radioligand were carried out for 30 min at 37 °C in 200 µl binding buffer supplemented with 10 µM glycine and 0.1% bovine serum albumin. Incubations were terminated by washing with ice-cold binding buffer for 10 s, followed by exposure to 1 + 0.5 ml lysis buffer (3 M acetic acid with 8 M urea and 2%, v/v, Nonidet P-40). Nonspecific binding, which amounted to 30–35% of total binding, was defined in the presence of 10 µM unlabelled phencyclidine (Research Biochemicals). Preliminary experiments indicated that the K_d for [¹²⁵I]iodo-MK801 was ~3 nM and this concentration of radioligand was used in three separate experiments to generate the data shown ($n=12$). *b*, Fourteen-day-old cultures were studied with respect to ⁴⁵Ca²⁺ uptake. Briefly, cultures were washed 3 times and preincubated for 5 min in the above buffer, without CaCl₂, before transfer to buffer containing 1.26 mM CaCl₂, 3 µCi ml⁻¹ ⁴⁵Ca²⁺ (Amersham) and 10 µM glycine, with or without 300 µM NMDA. Uptake was terminated by 3 brief wash steps in ice-cold buffer. The excitatory amino acid elicited a 2.9 ± 0.3-fold increase of ⁴⁵Ca²⁺ uptake when compared with basal uptake. The effect of 300 µM NMDA was blocked to 79 ± 6% by the presence of 10 mM (+)MK801. $N=6-9$ for ⁴⁵Ca²⁺ studies. *c*, NMDA-induced toxicity was studied in 20-day-old cultures. Aliquots of media were removed and saved. Wells with inverted coverslips were then immersed (with gentle shaking) for 30 min in remaining media, containing 10 µM glycine,



with or without 300 µM NMDA, after which the original 'conditioned' media were added back. On the following day (18–20 h later), the cells were stained with 4% Trypan blue dye and the number of dye-stained neurons and total neurons were counted as described⁶. Overall, approximately 60–70% of the neurons were scored as dead as a result of the NMDA exposure. (+)MK801 (10 µM) blocked 81 ± 15% of the NMDA effect in one triplicate experiment. Data shown in *c* are from 2 separate experiments in quadruplicate. All three panels are means with error bars representing s.e.m. Multiple comparisons were analysed using analysis of variance (ANOVA) and the Newman-Keuls test: * $P < 0.05$; ** $P < 0.01$.

FIG. 2 Reduction of ischaemic infarct volume after middle cerebral artery (MCA) occlusion in spontaneously hypertensive rats (SHR) pretreated with intracerebroventricular injections of NMDA-R1 antisense ODN (NMDAR1AS/c). The effect was compared with that of the non-competitive NMDA receptor blocker, (+)MK801. Bar graphs represent the four treatment groups: vehicle (H_2O ; $n=5$), antisense ODN ($n=6$), corresponding sense ODN ($n=6$), and (+)MK801 (given intravenously; $n=5$).

METHODS. Twenty-two adult male Wistar-Kyoto rats of the spontaneously hypertensive (SHR) strain weighing 300–350 g were used. Rats were anaesthetized with halothane (1.8–2.5% in 100% O_2) blown over the nose and stereotactically implanted with chronic 23-G guide cannulas (Plastics One) aimed at the lateral cerebral ventricle. The guide cannulas were secured to the skull with stainless steel screws and dental cement. After a 7-day recovery period, the ODN, NMDAR1AS/c, or the corresponding sense ODN was injected intracerebroventricularly, in a dose of 15 nmol in 5 μ l H_2O , or vehicle alone. Injections were made between 08:00 to 09:00 and 19:00 to 20:00 on days 1 and 2. On the third day, the animals were reanaesthetized with halothane and the MCA occluded as detailed¹⁴. In brief, thin-wall polyethylene (0.965 mm outer diameter) catheters were placed in both femoral arteries and one vein. One of the arterial catheters was connected to a Statham P23 db transducer for continuous monitoring of pulsatile and mean arterial blood pressure (MABP) and heart rate on a chart recorder (Grass Model 7D). Following instrumentation, rats were mounted in a Kopf stereotaxic apparatus. Body temperature was continuously monitored and maintained at 37 °C by a thermostatically controlled infrared lamp connected to a rectal probe. The MCA was exposed through a hole in the skull at a point rostral to the fusion of the zygomatic arch and squamous bone, and occluded just superior to the inferior cortical vein¹⁴. The same surgeon was responsible for production of all lesions. One group of rats received 1 mg kg⁻¹ (+)MK801 intravenously 30 min after the MCA occlusion. The MABP was comparable at the time of MCA occlusion among the four groups. But immediately on treatment with MK801, MABP was reduced significantly to ~70% of the preinjection level (from 140 ± 5 to 94 ± 6 mm Hg, $P < 0.001$), and MABP returned gradually to the previous level by the end of the experimental period (1 h after MCA occlusion). The cranial wound was covered with Gelform (Upjohn) and closed with sutures. The animals were removed from the stereotaxic frame, and the femoral arterial and venous catheters were capped. The anaesthesia was discontinued and the animal was returned to its cage. Twenty-four hours later, animals were reanaesthetized with halothane and killed by decapitation. Brains were removed, frozen, and serially sectioned at 20 μ m in the coronal plane in a cryostat microtome



at -20 °C. Slices were sampled every 200 μ m and stained with thionin. Each section was identified by reference to corresponding levels in the rat brain atlas²¹. The cross-sectional area was measured by tracing the lesion, which was defined by the almost complete loss of staining of Nissl substance, by computer-assisted image analysis (MCID; Imaging Research). The volume of the infarct was calculated as the product of cross-sectional area for all sections and distance between sections. The distribution of infarct lesion was also drawn by calculating the mean distance on the dorso-lateral surface of the brain between midline and the dorsal edge of the infarct area in the coronal plane. Infarct volumes are expressed as means $\pm$ s.e.m. in the bar graph. Statistical analyses involved analysis of variance (ANOVA) and the Newman-Keuls test. Differences were considered significant for $P < 0.05$ (*).

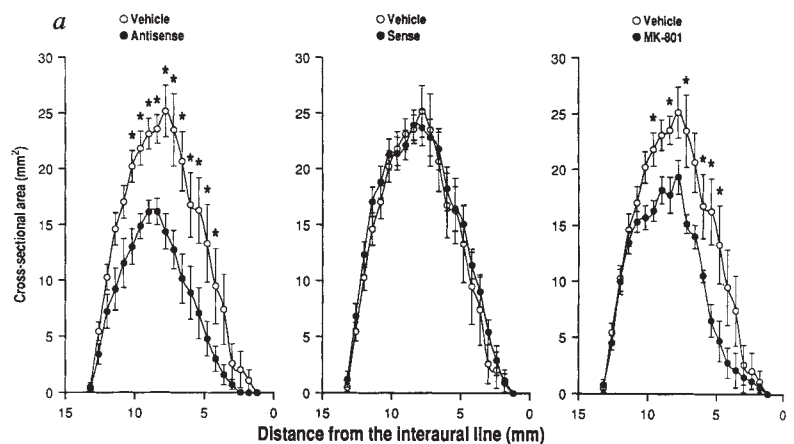
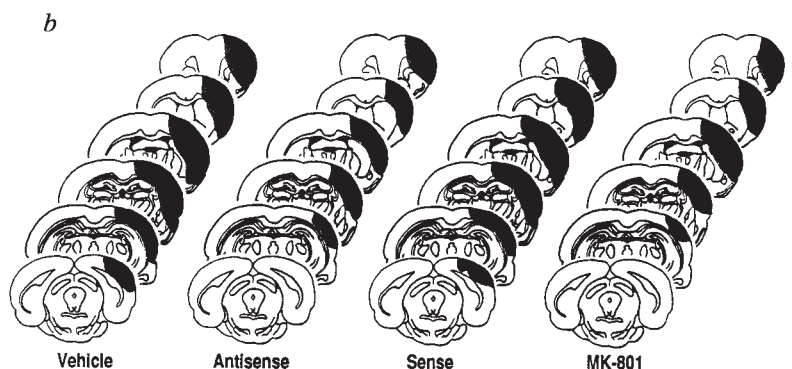


FIG. 3 Regional distribution of rat brain regions salvaged by the antisense ODN (NMDAR1AS/c) and by MK801 (animals are the same as for Fig. 2). Focal ischaemic infarctions were induced by occlusion of the middle cerebral artery (MCA) as described in Fig. 2 legend. The upper graphs (a) represent the cross-sectional area (mm²) at representative levels along the rostrocaudal axis as measured from the interaural line in animals treated with vehicle ($n=5$), antisense ODN ($n=6$), corresponding sense ODN ($n=6$), or (+)MK801 ($n=5$). Each data point is expressed as mean $\pm$ s.e.m. (* $P < 0.05$, analysis of variance and Newman-Keuls test). The lower panels (b) illustrate the superimposed distribution of the ischaemic infarct (dark stippled areas) of all animals in each group at representative levels (11.7, 9.7, 7.7, 5.7, 3.7 and 1.7 mm from the interaural line). Note that the distribution of the salvage by the treatment with antisense ODN or MK801 was contained within the penumbra.



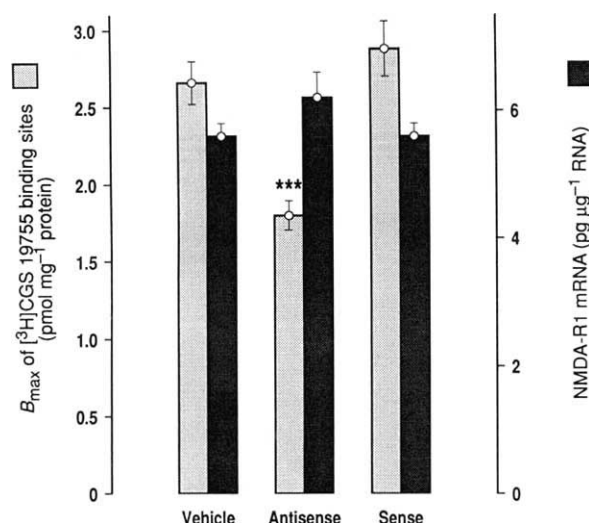


FIG. 4 Downregulation of [3 H]CGS 19755 labelled NMDA binding sites (lightly shaded bars), but not of NMDA-R1 mRNA (dark stippled bars), in spontaneously hypertensive rats (SHR) treated with antisense ODN (NMDAR1AS/c) *in vivo*.

METHODS. SHR received repeated intracerebroventricular injections of vehicle (H_2O ; $n=8$), antisense ODN (NMDAR1AS/c; $n=8$), or matching sense ODN ($n=8$) exactly as described in the legend for Fig. 2. They were killed ~54 h after the first injection. The cerebral cortex was rapidly dissected and most of it was subsequently used for radioreceptor binding experiments. A smaller part of the cortex, the frontal pole, was used for mRNA determination. For receptor binding with [3 H]CGS 19755 (New England Nuclear), we followed a centrifugation assay protocol, using crude synaptic membranes, essentially as described¹⁷. Briefly, our 250- μ l assay mixtures containing 0.1 mg protein were incubated for 30 min on ice. Specific binding was defined in the presence of 500 μ M L-glutamic acid (Sigma) or NMDA. Saturation curves (based on triplicate determinations using 6 concentrations of radioligand), constructed separately from each rat using computerized nonlinear least squares regression (Accufit and Accucomp), yielded individual estimates of binding site numbers (B_{max}) and binding affinities (dissociation constant, K_d). Group averages of these parameters were compared. The K_d values did not differ between groups. The lightly shaded bar graphs show means $\pm$ s.e.m. of the B_{max} values; $n=8$. mRNA was determined using a solution hybridization assay¹⁷. Briefly, total cellular RNA was extracted with guanidine-HCl/phenol from frontal poles. We used a 32 P-labelled 1,413-base riboprobe¹⁷ derived from the pN60 plasmid⁷, containing the rat NMDA-R1 cDNA (gift from S. Nakanishi). A sense transcript was used as calibration standard and an 18S riboprobe served to determine total RNA concentrations. Hybridization solutions were incubated at 75 $^{\circ}$ C for 4 h and then treated with RNase A (40 μ g ml⁻¹) and RNase T1 (2 μ g ml⁻¹); protected species were precipitated with trichloroacetic acid (TCA; 5%), collected onto glass microfibre filter paper, and finally washed 3 times with 5% TCA. Data are expressed as mean $\pm$ s.e.m. of pg NMDA-R1 mRNA per μ g total cellular RNA; $n=8$.

minipump), resulting in significant ($P<0.05$) reduction of NMDA binding site density.

These *in vivo* results are consistent with the view that NMDA receptors, and specifically the NMDA-R1 subunit⁷, may participate in the neurotoxicity elicited by focal cerebral ischaemia. Antisense ODNs, suppressing translation of specific neuronal molecules at any stage of development, may be useful tools for dissection of the participation of specific molecules in the expression of ischaemic lesions and conversely of neuroprotection. The experiments also establish the feasibility of an antisense oligonucleotide strategy for treatment of focal ischaemic infarctions. □

- Choi, D. W. *Neuron* **1**, 623–634 (1988).
- Park, K. K., Nehls, D. G., Graham, D. I., Teasdale, G. M. & McCulloch, J. *Ann. Neurol.* **24**, 543–551 (1988).
- Takizawa, S., Hogan, M. & Hakim, A. M. *J. cereb. Blood Flow Metab.* **11**, 786–793 (1991).
- Michaels, R. L. & Rothman, S. M. *J. Neurosci.* **10**, 283–292 (1990).
- Moryoshi, K., Masu, M., Ishii, T., Shigemoto, R. & Nakanishi, S. *Nature* **354**, 31–37 (1991).
- Nakanishi, N., Axel, R. & Schneider, N. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 8552–8556 (1992).
- Burnashev, N. et al. *Science* **257**, 1415–1417 (1992).
- Buchhalter, J. R. & Dichter, M. A. in *Neuromethods 23: Practical Cell Culture Techniques* (eds Boulton, A., Baker, G. & Walz, W.) 241–268 (Humana, Clifton, NJ, 1992).
- Ransom, R. W., Eng, W.-s., Burns, D., Gibson, R. E. & Solomon, H. F. *Life Sci.* **46**, 1103–1110 (1990).
- Wahlestedt, C., Merlo Pich, E., Koob, G. F., Yee, F. & Heilig, M. *Science* **259**, 528–531 (1993).
- Honore, T., Drejer, J., Nielsen, E. O. & Nielsen, M. *Biochem. Pharmac.* **38**, 3207–3212 (1989).
- Reis, D. J., Berger, S. B., Underwood, M. D. & Khayata, M. J. *J. cereb. Blood Flow Metab.* **11**, 810–818 (1991).
- Meldrum, B. & Garthwaite, J. *Trends pharmac. Sci.* **11**, 379–387 (1990).
- Gill, R., Andine, P., Hillered, L., Persson, L. & Hagberg, H. *J. cereb. Blood Flow Metab.* **12**, 371–379 (1992).
- Franklin, S. B., Elliott, K., Zhu, Y.-S., Wahlestedt, C. & Inturrisi, C. E. *Molec. Brain Res.* (in the press).
- Murphy, D. E., Hutchison, A. J., Hurt, S. D., Williams, M. & Silis, M. A. *Br. J. Pharmac.* **95**, 932–938 (1988).
- Toulme, J.-J. in *Antisense RNA and DNA* (ed. Murray, J. A. H.) 175–194 (Wiley-Liss, New York, 1992).
- Bottenstein, J. & Sato, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 514–517 (1979).
- Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates* 2nd edn (Academic, San Diego, 1986).

ACKNOWLEDGEMENTS. We thank L. Lyandvert for tissue culture assistance, S. Regunathan for help with *in vitro* studies, and M. Heilig for discussion. Supported in part by the National Heart and Lung Institute.

CFTR and outward rectifying chloride channels are distinct proteins with a regulatory relationship

Sherif E. Gabriel, Lane L. Clarke, Richard C. Boucher & M. Jackson Stutts

Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599–7020, USA

IN cystic fibrosis (CF), numerous epithelial cell functions are abnormal, including Cl^- conductance, sodium absorption, mucin sulphation and enzyme secretion^{1–4}. Although the CF gene product, the cystic fibrosis transmembrane conductance regulator (CFTR), functions as a small linear Cl^- channel^{5–8}, it is difficult to attribute such pleiotropic disease manifestations solely to a defect in Cl^- conductance. This has led to speculation that CFTR regulates the activity of other proteins. One possible example is the protein kinase A activation of outward rectifying Cl^- channels (ORCC), which is defective in membrane patches excised from CF cells^{9–16}. Whether CFTR regulates the activity of an independent anion channel is debatable, because ORCC occur exclusively in excised membrane patches and could be an excision-induced molecular derivative of CFTR. ‘Knockout’ mice that lack CFTR^{17,18} provide a means to define the relationship between CFTR and ORCC. Here we report that ORCC are present in CFTR(–/–) mouse nasal epithelial cells and thus cannot be a derivative of the CFTR molecule. Also ORCC were regulated by protein kinase A in membrane patches from normal but not CFTR(–/–) cells. These observations are the first, to our knowledge definitive demonstration that CFTR regulates the activity of another protein.

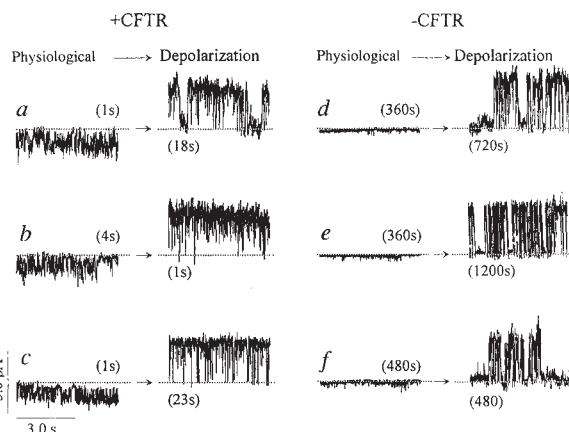
CFTR(–/–) cells lack full-length CFTR and the truncated message is expressed at only 1–2% of the normal message level¹⁷. To test the possibility that ORCC and CFTR arise from the same protein, we assessed the incidence of ORCC in cells from normal and CFTR(–/–) mice. ORCC were studied at depolarizing potentials in membrane patches excised from normal or CFTR(–/–) cells from a total of 21 mice (Table 1). The normal group consisted of 13 mice, of which 6 were CFTR(+/+) and 7 were CFTR(+/-); a total of 77 successful patches were formed. ORCC were identified in 28 of these patches, a frequency of

Received 2 December 1992; accepted 9 March 1993.

- Meldrum, B. S. *Clin. Sci.* **68**, 113–122 (1985).
- Rothman, S. & Olney, J. W. *Ann. Neurol.* **19**, 105–111 (1986).

FIG. 1 Outward rectifying chloride channels in membrane patches excised from normal and CFTR(-/-) cells into solutions containing PKA and ATP. Left, Representative traces from 3 different ORCC in excised patches from normal cells (a, b, and c) and recorded at physiological potentials and the same channels recorded at depolarizing potentials (+CFTR). Right, excised patches from 3 CFTR(-/-) cells (d, e, and f) recorded at physiological potentials and verification of the presence of the ORCC in the same patches recorded at depolarizing potentials (-CFTR). Closed states are indicated by the dotted line. Values above traces are the duration of the applied potential (in seconds) before the shown ORCC activity occurred.

METHODS. Mouse nasal epithelial cells were isolated from normal (CFTR(+/-; +/-)) or CFTR(-/-) mice as described previously¹⁸. Isolated cells were plated on collagen-coated dishes pretreated with 'Cell-tak' (5 $\mu\text{g cm}^{-2}$) and cultured at 37 °C in tissue culture media¹⁸. Patch clamp studies were done at 20 °C 2–5 days after plating. Pipette solution was composed of 150 mM NaCl, 10 mM TES (pH 7.2), 2 mM MgCl_2 and 1 mM CaCl_2 . Bath solutions were identical in composition to the pipette solution but also included 30 mM sucrose, 10 mM glucose, 2 mM Mg-ATP and 150 nM PKA-catalytic subunit. Excised membrane patches were held at 0 mV for 15–30 s immediately on excision. Patches were then exposed to physiological potentials either until channel activity occurred or for between 6 and 8 min (2 min at -40, 2 min at -60 and 2 to 4 min at -80). After the appearance of channel activity at physiological potentials, or after holding at physiological potentials for up to 8 min, depolarizing voltages were applied (for up to



30 min) to verify the presence of the ORCC. Patch currents were amplified by an Axopatch-1D patch-clamp amplifier and recorded on video cassette tape for later analysis. Data were low-pass filtered at 300 Hz and digitized at 3 kHz. All recordings, data analysis and selection of traces were completed blind to the genotype of the cells from which the ORCC were identified.

36.4% (28/77). Similarly, 8 CFTR(-/-) mice were studied and 37 excised inside-out patches were analysed. ORCC were observed in 14 patches, and the frequency of occurrence in CF cells 37.8% (14/37) was indistinguishable from that of normal cells. This dispels any molecular identity between the CFTR and ORCC as an explanation for previous reports of defective regulation of ORCC in CF.

We designed a simple protocol to evaluate ORCC regulation in the presence or absence of CFTR. Membrane patches from normal and CFTR(-/-) cells were excised at physiological voltage into solutions containing protein kinase A (PKA) and ATP (Fig. 1). Investigators were blinded to the cell genotype until all experiments and data analysis were complete. Two distinct patterns of ORCC activity were observed: (1) immediate

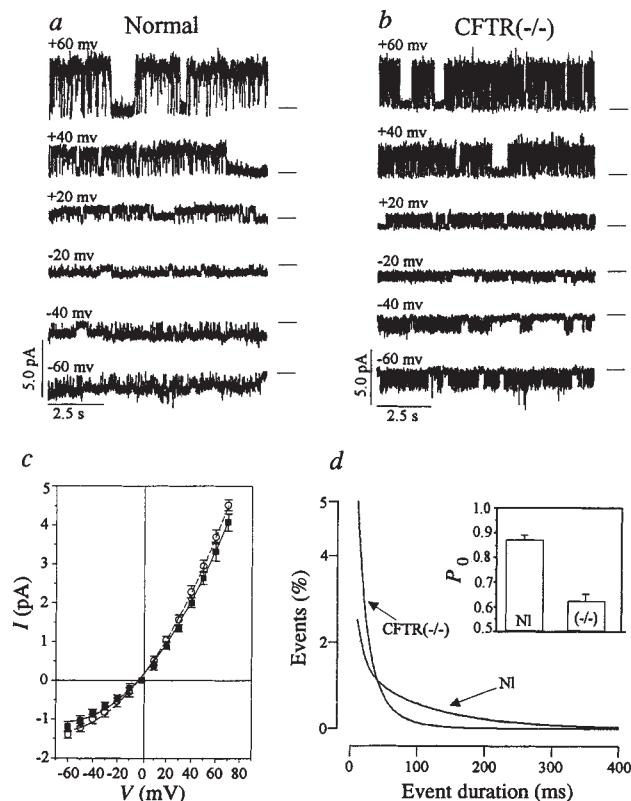


FIG. 2 Outward rectifying chloride channels in normal and CFTR(-/-) mouse nasal epithelial cells. Representative single channel traces are shown from nasal epithelial cells isolated from a normal CFTR(+/-) mouse (a), and from a CFTR(-/-) mouse (b); closed states are indicated by the dashed lines. c, Current versus voltage relationships are shown for ORCC from both normal (■) and CFTR(-/-) (○) cells; points represent mean and standard error measurements from 5 different ORCC from normal cells and 5 different ORCC from CFTR(-/-) cells and were determined from current-amplitude histograms. d, Mean open time (MOT) and closed time (MCT) distributions for normal (NI) and CFTR(-/-) cells were measured during bursts of channel activity at depolarizing potentials (11 normal files; with 861 events with a MOT=83.73 ms and 860 events with a MCT=6.75 ms, and 12 CFTR(-/-) files; with 2,102 events with a MOT=23.45 ms and 2,101 events with a MCT=9.42 ms); only MOT is shown. Inset, Open probability (P_o) of ORCC from normal and CFTR(-/-) cells was measured during bursts of channel activity of not shorter than 3.3 s and not interrupted by gaps of longer than 500 ms, as described previously³². In normal cells the ORCC activated by PKA and ATP studied at depolarizing potentials had a P_o of 0.87 ± 0.02 (5-membrane patches with a total burst time of 101.4 s) and P_o of 0.74 ± 0.06 (5-membrane patches with a total burst time of 66.5 s) at physiological potentials. In contrast, ORCC from CFTR(-/-) cells activated by strong depolarization had P_o values of 0.62 ± 0.03 (4-membrane patches with a total burst time of 86.3 s) and 0.55 ± 0.09 (4-membrane patches with a total burst time of 45 s) at depolarizing and physiological potentials, respectively; these values are statistically different from the comparable control values (depolarizing potentials, $P < 0.001$; physiological potentials, $P < 0.025$). In a separate study from this blind study we measured P_o of the ORCC excised from normal cells and activated by depolarization in the absence of exogenous PKA and ATP. The measured P_o at depolarizing potentials is indicative of a basal level of endogenous PKA activity ($P_o = 0.70 \pm 0.02$; 15-membrane patches with a total burst time of 340.7).

METHODS. Solutions and protocols used were similar to those used for Fig. 1. ORCC were identified by reversal potential (R_p), distinctive rectification properties, and the ability to voltage-activate the channel; furthermore, several channels were exposed to dilution of the bath solution with 330 mM sucrose and the appropriate shift in the R_p was detected both for normal and for CFTR(-/-) ORCC (normal $\Delta R_p = -15.3$ ($n=3$); CFTR(-/-) $\Delta R_p = -15.0$ ($n=3$)). All data used for kinetic analysis was filtered at 1 kHz and sampled at 5 kHz. P_o and MOT and MCT were determined by applying a 50% threshold of the amplitude histogram for each individual file.

TABLE 1 Incidence of anion channels in mouse nasal epithelial cells

	CFTR (+/+, +/-)	CFTR (-/-)
Excised, depolarizing (number of ORCC/patches)	28/77	14/37
On cell, $\uparrow$ cAMPi (number of CFTR-like/patches)	0/14	0/15

activity on excision at physiological voltage with similar immediate activity maintained at depolarizing voltage; and (2) inactive patches at physiological voltages but with ORCC activity appearing after prolonged depolarization. Seven of seventeen membrane patches excised from normal cells contained ORCC, of which five (71.4%) fit the pattern of immediate activation without depolarization (Fig. 1, left). All five patches also demonstrated rapid onset of ORCC activity within 30 seconds of the first exposure to +60 mV. A similar incidence of ORCC was found in membrane patches excised from CFTR(-/-) cells (8/20). But in contrast to normal patches, ORCC in CFTR(-/-) patches were not active (0 of 8 patches (0.00%)) at physiological potentials in the presence of PKA and ATP for 6–8 minutes (Fig. 1, right). Furthermore, depolarization-induced activation in CFTR(-/-) patches occurred only after additional hundreds of seconds (no ORCC appeared within the first 4 minutes of depolarization, and 3/8 patches required +60 mV, 3/8 patches required +80 mV, and 2/8 patches required +120 mV to demonstrate ORCC activity). ORCC in membrane patches excised from cells that do not contain CFTR are not activated by PKA-mediated phosphorylation.

Previous studies showing a relationship between CF mutations and regulation of ORCC by PKA have been questioned on several grounds. The ORCC is present probably only in excised membrane patches, its biophysical characteristics are not consistent with a role in epithelial Cl^- secretion, and it is present in many cells that do not secrete Cl^- (refs 19, 20). Moreover, ORCC are activated by many nonspecific mechanisms, including temperature, high salt concentration and, most characteristically, by strong depolarization²¹. By most parameters studied here, current amplitude, voltage activation, voltage dependence, current-voltage relationship and anion versus cation selectivity,

the ORCC in normal cells (Fig. 2a) were not different from ORCC in CFTR(-/-) cells (Fig. 2b). A cumulative I/V plot is shown for five different ORCC from normal cells and five different ORCC from CFTR(-/-) cells (Fig. 2c). Our blinded protocol with membrane patches from normal and CFTR(-/-) mouse nasal cells excised directly into a bath containing PKA and ATP at physiological voltage eliminated subjective scoring of activation and minimized the possibility of nonspecific activation. As already described (Fig. 1), a regulatory relationship between the CFTR and ORCC is unequivocal. In addition, although a detailed kinetic analysis was not the objective of this design, it was obvious that the ORCC activated by PKA and ATP in excised patches from normal cells showed a greater open probability (P_o), an increased mean open time and decreased mean closed time than did the ORCC from CFTR(-/-) cells (Fig. 2d). The P_o of normal ORCC in the absence of PKA and ATP was intermediate between that of CF and normal ORCC in the presence of PKA and ATP (see Fig. 2 legend), consistent with a basal level of endogenous PKA activity. That PKA-induced increases in ORCC P_o depend on the presence of CFTR is further evidence of a regulatory relationship between these two proteins.

CFTR studied in heterologous expression systems or in high-expressing tissues is seen in cell-attached membrane patches as small linear Cl^- channels^{5,7,22,23}. We used similar techniques in a second blinded protocol designed to quantify the channels present in cell-attached membrane patches of normal and CFTR(-/-) cells exposed to forskolin (10 mM) and cpt(8-(4-chlorophenylthio))-cAMP (500 mM). However, we observed no small linear Cl^- channels (0/13 patches from 4 control mice and 0/15 patches from 4 CFTR(-/-) mice) (Table 1). Because we detected small linear Cl^- channels²⁴, in several nonepithelial cell lines in which CFTR may be expressed at very low levels²⁵, this absence of Cl^- channels is, therefore, not probably due to a sensitivity problem. Given recent reports that CFTR expression and function as a Cl^- conductance in some epithelia requires polarization^{26–28}, we tested the ability of forskolin to activate a Cl^- conductance in mouse nasal epithelial cells grown to confluence on permeable supports (Fig. 3a). In contrast to polarized preparations, cpt-cAMP and forskolin failed to activate whole cell Cl^- conductance in isolated mouse nasal cells grown subconfluently on impermeable supports, even though the same conditions activated Cl^- conductance in both 3T3-fibroblasts express-

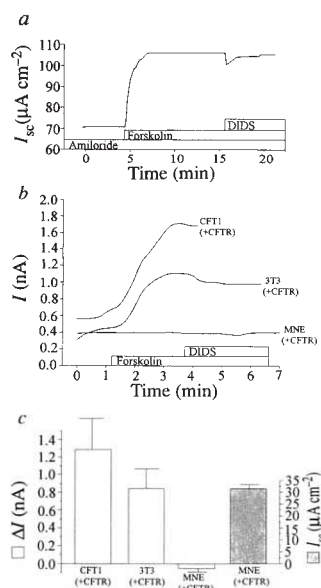


FIG. 3 Cyclic AMP-stimulated Cl^- currents in confluent polarized epithelia and in nonpolarized disaggregated cells. **a**, A representative short-circuit current (I_{sc}) trace from a confluent normal mouse nasal epithelium mounted in a conventional Ussing chamber. **b**, Representative whole-cell outward current recordings from a CFT1 human airway epithelial cell stably expressing human CFTR (CFT1(+CFTR)), an NIH3T3 fibroblast cell stably expressing human CFTR (3T3(+CFTR)), and from a normal mouse nasal epithelial cell (MNE(+CFTR)). **c**, Cumulative data of the mean change in current after the addition of forskolin to isolated mouse cells MNE(+CFTR) ($n=7$), 3T3(+CFTR) cells ($n=8$), CFT1(+CFTR) cells ($n=9$), and a confluent mouse nasal epithelium MNE(+CFTR) ($n=7$); values are mean $\pm$ s.e.

METHODS. Conventional whole-cell patch-clamp protocols were used. The pipette solution was 40 mM Tris-Cl, 100 mM Tris-gluconate, 1 mM EGTA, 2 mM MgCl_2 , 5 mM TES (pH 7.2) and 2 mM ATP. The bath solution contained 150 mM Tris-Cl or NaCl, 2 mM MgCl_2 , 1 mM CaCl_2 , 5 mM TES (pH 7.2) and 30 mM sucrose. Cl^- currents were identified by distinctive reversal potentials. Whole-cell currents were measured at a holding potential of -40 mV and stepped to +60 mV every 10 s for 0.45 s. Mouse cells used for whole-cell studies were isolated, cultured and plated to subconfluence on collagen-coated plastic dishes as described in Fig. 1. NIH3T3-fibroblasts and recombinant CFT1 airway cells were cultured and plated on plastic 35-mm dishes for patch-clamp studies as described previously²⁴. Culture and experimental conditions for mouse nasal epithelial cells in Ussing chambers have also been described¹⁸. Drug additions to Ussing preparations were pretreated with amiloride (10^{-4} M), followed by additions to achieve a final concentration of 10^{-5} M forskolin, and 5×10^{-4} M DIDS (4,4'-diisothiocyanato-2,2'-stilbene disulphonic acid).

ing human CFTR, and CFT1(+CFTR) cells, a human CF airway epithelial cell expressing recombinant human CFTR (Fig. 3b). From these data it is evident that cAMP stimulates Cl^- currents in confluent polarized mouse nasal epithelia or in high-expressing heterologous cells, but not in nonpolarized mouse nasal epithelial cells (Fig. 3c). Thus, the lack of cAMP-activated Cl^- channels in CFTR(+/-; +/-) cells cultured on impermeable supports would seem to relate to lack of polarization, consistent with studies of other cells expressing CFTR^{26,27} or, equivalently, to low expression of endogenous murine CFTR. Intriguingly, nonpolarized mouse nasal epithelial cells containing CFTR still demonstrate regulation of ORCC even though we detected no CFTR-like channels or cAMP-activated whole-cell currents. This finding suggests that a low copy number of CFTR in the plasma membrane may be sufficient to regulate the activity of another distinct protein, the ORCC. Alternatively, the function of CFTR as a Cl^- channel at a site other than the plasma membrane, such as Golgi or endosomes^{29,30}, could affect the

regulation of ORCC.

Because the CFTR(-/-) mouse is completely devoid of full-length CFTR, we have been able to determine unequivocally that there is no molecular identity between the CFTR molecule and the ORCC. This lays to rest any suspicion that ORCC are derivatives of the CFTR molecule, modified by excision and strong depolarization. Our results go further than previous studies to demonstrate that CFTR-associated PKA activation of ORCC results in increased open probability and altered channel kinetics. Because CFTR cannot be the ORCC, this difference in PKA-mediated activation of ORCC must represent an effect of CFTR on another protein. Owing to the undefined role of ORCC in cell physiology³¹, the functional significance of this regulation remains unknown. Nonetheless, the realization that CFTR affects the function of another protein indicates that some processes affected in CF disease may lack normal regulation by CFTR. □

Received 2 February; accepted 17 March 1993.

1. Knowles, M. R. *et al. Science* **221**, 1067-1070 (1983).
2. Frizzell, R. A. *Trends Neurosci.* **10**, 190-193 (1987).
3. Cheng, P.-W., Boat, T. F., Cranfill, K., Yankaskas, J. R. & Boucher, R. C. *J. clin. Invest.* **84**, 68-72 (1989).
4. Quinton, P. M. *FASEB J.* **4**, 2709-2717 (1990).
5. Kartner, N. *et al. Cell* **64**, 681-691 (1991).
6. Anderson, M. P., Rich, D. P., Gregory, R. J., Smith, A. E. & Welsh, M. J. *Science* **251**, 679-682 (1991).
7. Berger, H. B. *et al. J. clin. Invest.* **88**, 1422-1431 (1991).
8. Bear, C. E. *et al. Cell* **68**, 809-818 (1992).
9. Egan, M. *et al. Nature* **358**, 581-584 (1992).
10. Frizzell, R. A., Reckemmer, G. & Shoemaker, R. L. *Science* **233**, 558-560 (1986).
11. Li, M. *et al. Science* **244**, 1353-1356 (1989).
12. Chen, J. H., Schulman, H. & Gardner, P. *Science* **243**, 657-660 (1989).
13. Schoumacher, R. A. *et al. Nature* **330**, 752-754 (1987).
14. Hwang, T. C. *et al. Science* **244**, 1351-1353 (1989).
15. de Jonge, H. R., van den Bergh, N., Tilly, B. C., Kansan, M. & Bijman, J. *Biochem. Soc. Trans.* **17**, 816-818 (1989).
16. Jetten, A. M., Yankaskas, J. R., Stutts, M. J., Willumsen, N. J. & Boucher, R. C. *Science* **244**, 1472-1475 (1989).

17. Snouwaert, J. *et al. Science* **257**, 1083-1088 (1992).
18. Clarke, L. L. *et al. Science* **257**, 1125-1128 (1992).
19. Ward, C. L., Krouse, M. E., Gruenert, D. C., Kopito, R. R. & Wine, J. J. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5277-5281 (1991).
20. Anderson, M. P., Sheppard, D. N., Berger, H. A. & Welsh, M. J. *Am. J. Physiol.* **263**, L1-L14 (1992).
21. Tabcharani, J. A. & Hanrahan, J. W. *Am. J. Physiol.* **261**, 992-999 (1991).
22. Tabcharani, J. A., Low, W., Elie, D. & Hanrahan, J. W. *FEBS Lett.* **270**, 157-164 (1990).
23. Tabcharani, J. A., Chang, X. B., Riordan, J. R. & Hanrahan, J. W. *Nature* **352**, 628-631 (1991).
24. Gabriel, S. E., Price, E. M., Boucher, R. C. & Stutts, M. L. *Am. J. Physiol.* **263**, C708-C713 (1992).
25. Krauss, R. D. *et al. EMBO J.* **11**, 875-883 (1992).
26. Morris, A. P., Cunningham, S. A., Benos, D. J. & Frizzell, R. A. *J. biol. Chem.* **267**, 5575-5583 (1992).
27. Morris, A. P., Cunningham, S. A., Tousson, A., Benos, D. J. & Grizzell, R. A. *Pediatr. Pulmonol.* **8**, suppl. 247 (1992).
28. Haws, C., Krouse, M. E., Xia, Y., Gruenert, D. C. & Wine, J. J. *Pediatr. Pulmonol.* **8**, suppl. 253 (1992).
29. Bradbury, N. A. *et al. Science* **256**, 530-532 (1992).
30. Luckacs, G. L. *et al. J. biol. Chem.* **267**, 14568-14572 (1992).
31. Fischer, H. *et al. Pfluegers Arch.* (in the press).
32. Stutts, M. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 1621-1625 (1992).

Neurotrophins promote motor neuron survival and are present in embryonic limb bud

Christopher E. Henderson*, William Camu*, Clément Mettling*, Annie Gouin*, Kristian Poulsen†, Mona Karihaloo‡, Janette Rullamas†, Tony Evans†, Stephen B. McMahon‡, Mark P. Armanini§, Lucy Berkemeier§, Heidi S. Phillips§ & Arnon Rosenthal§

* Centre de Recherche de Biochimie Macromoléculaire, UPR 9008 du CNRS/U249 de l'INSERM, B.P. 5051, 34033 Montpellier, France

† Department of Cell Analysis and § Department of Neuroscience, Genentech Inc., South San Francisco, California 94080, USA

‡ Department of Physiology, UMDS (St Thomas' Campus), Lambeth Palace Road, London SE1 7EH, UK

Embryonic spinal motor neurons are thought to depend for survival on unidentified factors secreted both by their peripheral targets and by cells within the central nervous system¹. The neurotrophins are a family of polypeptides required for survival of discrete central and peripheral neuronal populations *in vivo* and *in vitro*^{2,3}. In spite of their ability to reduce motor neuron death *in vivo*⁴⁻⁶, the known neurotrophins have been thought to be without direct effect on motor neurons⁷⁻¹⁰. Here we show that picomolar concentrations of three of them, brain-derived neurotrophic factor, neurotrophin-3 and neurotrophin-5, can prevent the death of cultured embryonic rat spinal motor neurons. Furthermore, messenger RNA coding for neurotrophins is present at appropriate stages in spinal cord

and limb bud, and mRNA for their receptors is found in motor neurons. These neurotrophins may therefore be physiological motor neuron growth factors.

Enriched cultures of motor neurons were prepared from ventral spinal cords of 15-day-old rat embryos, an age at which motor neuron cell death is occurring in the spinal cord, using a novel density gradient-panning method (see legend to Fig. 1; ref. 11). Many of the cells were large (cell body diameter $35 \pm 2 \mu\text{m}$, longest neurite $1,300 \pm 100 \mu\text{m}$, mean $\pm$ s.e.m., $n = 16$), multipolar, with tapering thicker processes and fine axon-like neurites (Fig. 1a-c), and expressed p75 low-affinity neurotrophin receptor (Fig. 1d). In addition, over 95% of the large neurons were stained by antisera to L14, an intracellular lectin specifically synthesized by motor neurons in the spinal cord¹² (Fig. 1e), and to Islet-1 homeoprotein¹³, one of the earliest motor neuron markers (Fig. 1f, g). To confirm that cultures were indeed enriched for motor neurons, neurons were retrogradely labelled *in situ* by injection of 1% tetramethylrhodamine dextran into limb buds¹⁴ and then purified. Fluorescent cells increased from 0.3% in total ventral spinal cord dissociates to 21% in cultures purified by metrizamide-panning. In both E14 and E15 cultures, smaller neurons, diameter $18.4 \pm 0.9 \mu\text{m}$, longest neurite $230 \pm 30 \mu\text{m}$, no tapering dendrites) also developed in variable proportions (range 6-80%; mean $50 \pm 9\%$ in 9 experiments). They stained for both p75 and L14, but not Islet-1. Large and small neurons both survive better in the presence of neurotrophins. Because the small neurons did not stain for Islet-1, the two neuronal populations were analysed separately and the small neurons excluded from motor neuron counts given in the figures.

Panned E15 motor neurons were seeded at 220 cells per cm^2 (on average 110 large motor neurons per cm^2) and viability was

evaluated after 15 h in culture (see legend to Fig. 2). At this time, the average number of large motor neurons was 56 ± 8 per cm^2 . The remainder were small neurons (56 ± 10 per cm^2) or cells that did not recover from the purification procedure. Of the large neurons, only 10 ± 2 neurons per cm^2 survived after 3 days in basal culture medium (Fig. 2). As expected^{8,9,14}, of the motor neurons that originally developed in culture, muscle extract ($25 \mu\text{g ml}^{-1}$ protein) supported 60–90%, ciliary neurotrophic factor (5 ng ml^{-1}) supported 26–28% and cholinergic development factor–leukaemia inhibitory factor (5 ng ml^{-1}) supported 28–33% in different experiments using E14 and E15 neurons. Basic fibroblast growth factor (10 ng ml^{-1}), transforming growth factor- $\beta 1$ (1 ng ml^{-1}) and platelet-derived growth factor (5 ng ml^{-1}) had only small effects (Fig. 2a). Surprisingly, given their reported lack of activity on chicken motor neurons^{8,10}, brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3) and neurotrophin-5 (NT-5) (100 pg ml^{-1} each) all had strong survival-promoting activity in our rat cultures, whereas nerve growth factor (NGF; 100 pg ml^{-1}) was without effect (Fig. 2a). Relative to optimal concentrations of muscle extract, individual neurotrophins supported the survival of $96 \pm 5\%$ (BDNF; $n = 4$), $79 \pm 6\%$ (NT-3; $n = 4$) and $106 \pm 11\%$ (NT-5; $n = 5$) of the large motor neurons, raising the possibility that the three neurotrophins support survival of the same neuronal population. This was confirmed by the observation that no further increase in motor neuron survival occurred when combinations of neurotrophins were used (Fig. 2b).

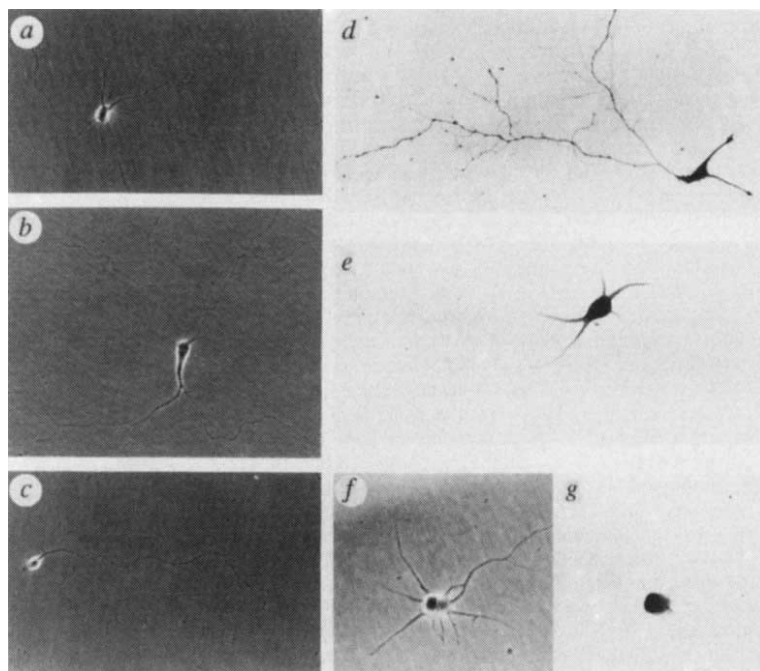
BDNF, NT-3 and NT-5 were extremely potent as survival factors for motor neurons (Fig. 2c): half-maximal activity was obtained at nominal concentrations ranging from 1 – 25 pg ml^{-1}

(10^{-13} – 10^{-12} M) in different experiments with different neurotrophins. Concentrations of 100 pg ml^{-1} ($4 \times 10^{-12} \text{ M}$) and above were saturating (Fig. 2c). After 6 days in 100 pg ml^{-1} BDNF, 50% of large motor neurons remained healthy, even without further addition of neurotrophin or replacement of medium (data not shown). In addition, we observed fourfold increases in choline acetyltransferase (ChAT) activity in cultures of metrizamide-enriched E15 ventral neurons in response to higher concentrations ($>100 \text{ pg ml}^{-1}$) of NT-5, BDNF and NT-3, whereas NGF had no significant effect at concentrations up to 20 ng ml^{-1} (Fig. 2d). Such increases may reflect enhanced survival of motor neurons and/or elevated ChAT activity per cell.

To avoid the possibility that panning might selectively enrich for neurotrophin-responsive subpopulations, ventral neurons from E14 embryos were also analysed after fractionation on a density gradient only. Over 95% of the large neurons in these cultures also stained for p75, L14 and Islet-1 and their response was identical to those of the panned neurons (not shown). The survival of the smaller L14- and p75-positive neurons described above was also higher in the presence of BDNF, NT-3 and NT-5 but not in the presence of NGF. At saturating concentrations (100 pg ml^{-1}), BDNF, NT-3 and NT-5 each increased the survival of these neurons on average 1.6-fold. Because the small neurons express both p75 and L14, we believe that they may correspond to the immature motor neurons seen by others in chicken⁸. Nevertheless, to exclude the possibility that they might indirectly mediate actions of neurotrophins on large motor neurons, neurons were prepared either by fluorescence-activated cell sorting of retrogradely labelled motor neurons or by a

FIG. 1 Motor neurons purified by the metrizamide-panning procedure. *a–c*, Phase-contrast micrographs of large neurons grown for 3 days in 100 pg ml^{-1} BDNF (*a*), 100 pg ml^{-1} NT-5 (*b*) and $25 \mu\text{g ml}^{-1}$ muscle extract (*c*). *d–g*, Identification of large neurons as motor neurons after 3 days in 100 pg ml^{-1} BDNF, by immunoperoxidase staining using antibodies to p75 neurotrophin receptor (*d*), L14 lectin (*e*) and Islet-1 homeoprotein (*g*). The phase-contrast image of *g* is shown in *f*. Scale bars, $100 \mu\text{m}$ shown in *c* for *a–d* and in *g* for *e–g*.

METHODS. Ventral spinal cords were dissected from E15 rat embryos, dissociated and loaded onto a metrizamide cushion as described^{9,11}. The cells were then incubated for 1 h in Petri dishes prepared for panning as described^{9,11} except that the first layer of anti-mouse IgG was incubated for 2 h at room temperature in PBS with $10 \mu\text{g}$ of 192 antibody (Boehringer–Mannheim). This antibody recognizes the p75 low-affinity neurotrophin receptor, which is only expressed by motor neurons within the E15 rat spinal cord. Tightly bound cells were removed by agitation with $10 \mu\text{g}$ 192 antibody in 3 ml L15 medium and seeded at a density of 200 cells per cm^2 in 35-mm tissue culture dishes²⁵ coated with polyornithine ($500 \mu\text{g ml}^{-1}$ in borate, pH 8.3) and laminin ($20 \mu\text{g ml}^{-1}$ in L15-bicarbonate). After 3 days in culture, cells to be stained were fixed for 15 min in ice-cold 4% paraformaldehyde (p75, Islet-1) or 2% paraformaldehyde, 0.1% glutaraldehyde (L14) in PBS. After permeabilization with 0.1% Triton X100 and blocking of non-specific sites, they were incubated successively with primary antibody ($5 \mu\text{g ml}^{-1}$ IgG-192 for p75, 1:3,000 anti-L14 rabbit serum, 1:1,000, anti-Islet 1 rabbit serum), appropriate biotinylated secondary antibody (1:100) and streptavidin-peroxidase (1:200). Peroxidase was revealed using diaminobenzidine, followed in *d* by silver enhancement (Amersham). Controls in which primary antibody was omitted or replaced by an irrelevant serum, in which secondary antibody was omitted or in which cells were not permeabilized (L14) gave no staining (not shown). Roughly 5,000 Islet-1-positive neurons per spinal cord survived the metrizamide and panning steps and put out neurites in culture.



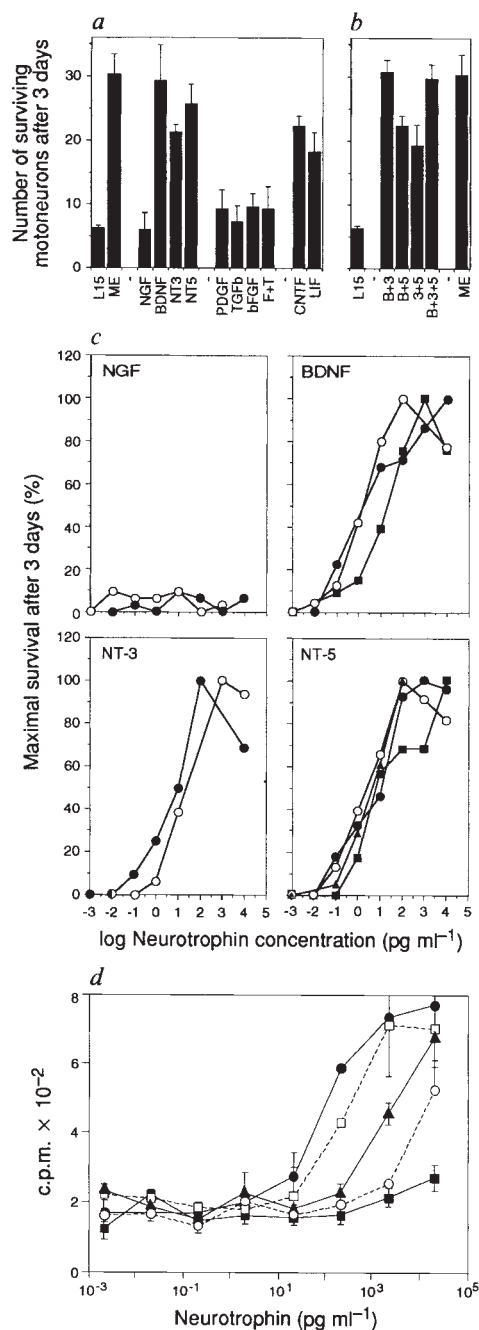
modification of the panning method giving >90% large Islet-1-positive motor neurons but lower yields. After 2 days in culture, survival was enhanced 4.3-fold (sorted neurons) or 4.1-fold (panned neurons) in the presence of 100 pg ml⁻¹ of each of the neurotrophins except NGF, suggesting that neurotrophin action is direct.

In other systems, BDNF, NT-3 and NT-5 mediate their biological effects through the tyrosine kinase receptors *trkB* and *trkC*^{10,15-17}. To ascertain the presence of *trk* receptors on cultured motor neurons, RNA from purified motor neurons was reverse-transcribed and then amplified in a polymerase chain reaction (PCR) using *trkA*-, *trkB*- or *trkC*-specific primers (RT-PCR). Motor neurons were found to contain mRNA for *trkB* and *trkC*, but only very low levels of *trkA* (Fig. 3A). This was confirmed by *in situ* hybridization on sections of embryonic mouse spinal

cord from a comparable developmental stage (Fig. 3B): motor neurons *in vivo* contain transcripts of the *trkB* and *trkC*, but not *trkA*, receptor genes. Similar findings have been reported for *trkA* and *trkB*¹⁸⁻²¹. These observations are consistent with a direct effect of neurotrophins on motor neurons. To determine whether adult motor neurons retain expression of *trkB* and *trkC*, the localization of *trk* mRNAs to adult rat motor neurons was analysed by combined retrograde fluorogold (FG) tracing and *in situ* hybridization. In agreement with the observation that adult motor neurons can retrogradely transport BDNF and NT-3, but not NGF²², *in situ* hybridization revealed that 22/23 FG-identified motor neurons expressed *trkB* mRNA, 23/28 expressed signal for *trkC* mRNA, whereas none of the cells that were FG positive expressed a clear signal for *trkA* mRNA (0/37) (Fig. 3B). These observations indicate that motor neurons in

FIG. 2 Effects of neurotrophins and other growth factors on motor neurons from 15-day rat embryos. *a*, Number of motor neurons surviving per cm² after culture for 3 days in: L15, complete culture medium; ME, 25 µg ml⁻¹ neonatal chicken muscle extract; NGF, BDNF, NT-3, NT-5, 100 pg ml⁻¹ human neurotrophins; PDGF, 5 ng ml⁻¹ human platelet-derived growth factor; TGFβ, 1 ng ml⁻¹ human transforming growth factor-β1; bFGF, 10 ng ml⁻¹ human basic fibroblast growth factor; F+T, combination of same concentrations of TGFβ1 and bFGF; CNTF, 5 ng ml⁻¹ rat ciliary neurotrophic factor; LIF, 5 ng ml⁻¹ human cholinergic development factor-leukaemia inhibitory factor. Results are expressed as mean ± s.e.m. of three different culture dishes in the same experiment. *b*, Lack of additivity of neurotrophin effects on motor neuron survival. All neurotrophins were used at 100 pg ml⁻¹. B, BDNF; 3, NT-3; 5, NT-5. *c*, Dose-response curves for neurotrophin enhancement of motor neuron survival. Each curve represents an individual experiment using E15 (closed symbols) or E14 (open symbols) motor neurons. The value in basal medium in each experiment was taken as 0% and the maximal value attained in each curve as 100%. Points are means of duplicate dishes; error bars were of the same order as those in *a* and *b*. *d*, Effect of NGF (■), BDNF (□), NT-3 (○), NT-5 (●) and CDF-LIF (▲) on choline acetyltransferase activity of enriched motor neuron cultures.

METHODS. Motor neurons from E15 or E14 rat embryos were seeded as described in the legend to Fig. 1 in the presence of the reported optimal concentrations of growth factors^{8,14,26} or with the indicated concentrations of neurotrophins. All recombinant factors were made at Genentech, except bFGF and PDGF (R & D Systems). Chicken muscle extract, prepared as described⁹, was more active than rat extracts and therefore was used as a positive control. Neurons were seeded at nominal densities of 200–500 cells per cm². Cell counts were done within an identified region (15% of the total surface area) in the centre of the dish; a total of up to 100 large neurons were counted in each condition. The highest survival value (the number of cells with neurites >4 cell diameters in length) after 15 h in a given experiment represented the number of viable motor neurons that could potentially be kept alive, and was used as 100% (in 6 experiments, this was 28 ± 4% of the value expected if all seeded cells had attached and developed neurites). BDNF, NT-3 and NT-5 at 100 ng ml⁻¹ each maintained on average 80 ± 6% of the motor neurons that initially developed in culture (range 45–103% in 10 experiments using all three neurotrophins and both E14 and E15 motor neurons). The smaller neurons did not contribute to the 100% value. Even at 15 h, motor neuron numbers in basal medium and NGF were lower than in the presence of other neurotrophins or muscle extract, although plating efficiencies after 4 h were indistinguishable, suggesting that death of some neurons occurs during the neurite outgrowth phase. Most motor neurons that were lost in L15 died by day 2; the other 19 ± 3% (mean ± s.e.m.; *n* = 6) survived in basal medium for 3 days and longer without further trophic support; similar findings have been reported using sorted rat motor neurons¹⁴. Motor neurons surviving in L15 were labelled by all the antibodies described and had neurite lengths comparable to those grown with neurotrophins. They may represent motor neurons that have not yet acquired trophic dependence²⁷. For ChAT activity, ventral spinal neurons from E15 rat embryos were enriched using metrizamide cushions and seeded at a density of 5,000 cells per well in 96-well dishes, in the presence of the indicated concentrations of neurotrophins. Three days later, the medium was removed and ChAT activity assayed in triplicate wells by the method of Fonnum²⁸. The apparent upward trend at high NGF concentrations in this experiment was not significant, even at 100 ng ml⁻¹ (not shown). The stimulation of ChAT activity by 20% (v/v) chicken muscle extract was > double that obtained using any of the factors alone at the concentrations indicated (not shown).



the adult rat lumbar region express both *trkB* and *trkC* *in vivo* and raise the possibility that neurotrophins could be used for the treatment of motor neuron diseases in adults.

A necessary condition for a physiological motor neuron growth factor is that it should be present in the environment of the motor neuron during the period of naturally occurring cell death. By northern blot analysis, mRNA for BDNF could be detected at low levels in both spinal cord and limb bud of E15 embryos, that for NT-3 was abundant in both spinal cord and limb bud, and NT-5 mRNA could be detected in limb but not spinal cord (Fig. 4A). High levels of NT-3 in spinal cord using northern blots have been reported^{23,24}. *In situ* hybridization on sections of E15 rat embryos was used to extend these findings (Fig. 4B). BDNF was expressed patchily in spinal cord and some muscles, NT-5 signal was detectable in developing skin

of the limb bud, and NT-3 was strongly present in both motor neurons and muscle, as already reported for NT-3 in mouse^{19,24}. BDNF and NT-3 are therefore likely to be present in the vicinity of the motor neuron during the cell death period and their dissimilar distribution in embryonic muscle may indicate distinct functions.

The unexpected biological activities and mRNA localization studies reported here are consistent with the hypothesis that members of the neurotrophin family are involved in regulation of motor neuron survival during embryonic development. Although all three neurotrophins were found to promote the survival of the same motor neuron population in our cultures, it is possible that at other developmental stages individual neurotrophins may play distinct roles. In particular, central and peripheral influences on motor neuron survival may be mediated

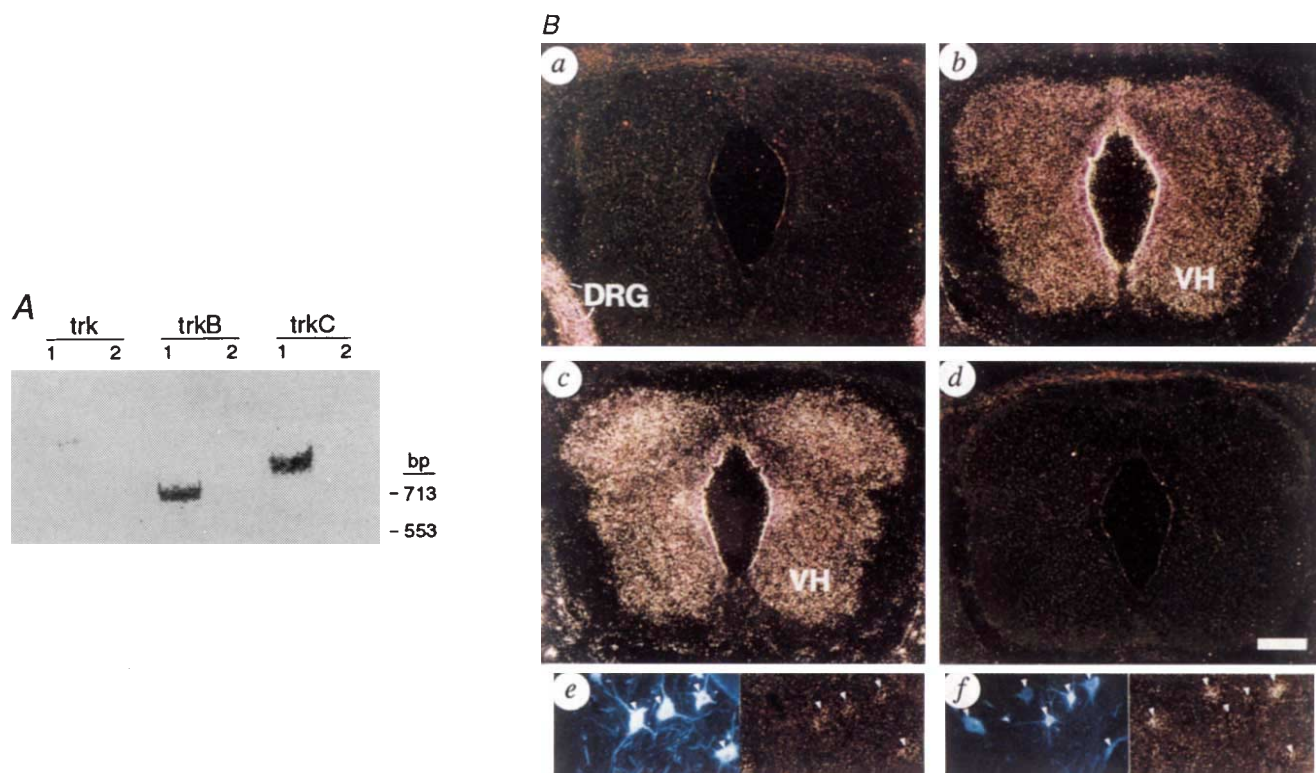


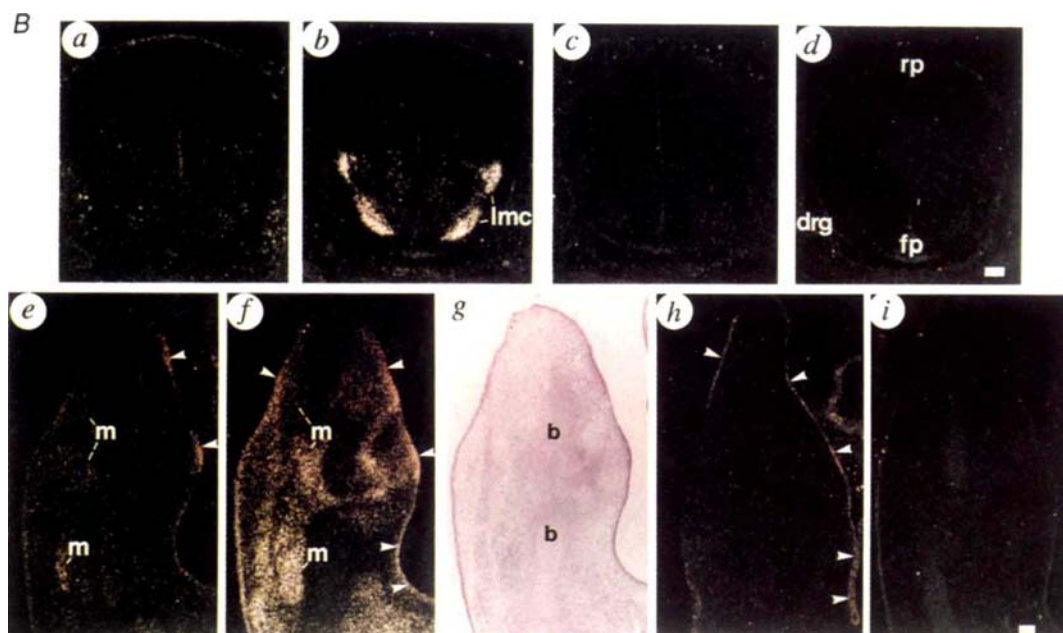
FIG. 3 Messenger RNA for neurotrophin receptors is present in motor neurons. **A**, Analysis of mRNA for *trkA*, *trkB* and *trkC* in cultured motor neurons using RT-PCR. Lanes 1, RT-PCR: cDNAs were used as templates for the PCR reaction; 2, PCR: RNA samples were used without reverse transcription as templates in the PCR reaction, as a control for contamination by genomic DNA. **B**, Expression of *trkA*, *trkB*, and *trkC* mRNAs in developing and adult rodent spinal cord. **a**, Antisense probe to *trkA*, **b**, **e**, Antisense probe to *trkB*. **c**, **f**, Antisense probe to *trkC*. **d**, Sense probe to NT-3 (control). **a–d**, Sections of E13.5 mouse cervical spinal cord reveal that hybridization for *trkB* and *trkC* is prominent in ventral horn (VH) as well as other regions of the developing grey matter. Although *trkA* is prominently expressed in dorsal root ganglia (DRG), no hybridization is seen within spinal cord. Scale bar, 100 μ m. **e**, **f**, Retrograde identification of individual adult rat lumbar motor neurons (arrowheads) with fluorogold and matching autoradiograms after *in situ* hybridization for *trkB* or *trkC*.

METHODS. **A**, ~100 ng of poly(A)⁺-RNA from motor neurons purified by metrizamide-panning was incubated with 20 μ g RNase-free deoxyribonuclease (Worthington) for 1 h at 37 $^{\circ}$ C. The sample was used in a reverse transcription reaction and was then divided into 3 PCR tubes (reaction volume 25 μ l) each containing 10 pmol *trkA* (amino acids 34–42), *trkB* (amino acids 395–403) or *trkC* (amino acids 2–11) specific sense primer, 10 pmol *trkA* (amino acids 327–335), *trkB* (amino acids 626–635) or *trkC* (amino acids 251–260) specific antisense primer. The PCR reaction was performed

for 30 cycles of 95 $^{\circ}$ C for 1 min, 60 $^{\circ}$ C for 1 min, 72 $^{\circ}$ C for 1 min as previously described²⁹. Samples were then run on a 6% polyacrylamide gel and exposed to a Fujix BAS 2000 image analyser. Samples containing PCR primers and RT mix without reverse transcriptase were negative in the PCR reaction. The *trkA* primers amplified cDNA from mouse E13 trigeminal neurons and the PCR primers for the three receptors amplified 10 pg of cDNAs encoding the corresponding receptors (not shown). The identity of the PCR products was confirmed by electrophoresing nonradioactive amplified fragments into a 1% agarose gel, blotting them onto a nylon membrane and hybridizing at high stringency with ³²P-labelled synthetic oligonucleotides derived from unique regions corresponding to amino acids 542–562 of rat *trkB* and amino acids 28–48 of rat *trkC*. **B**, Embryos were fixed by immersion in cold 4% formaldehyde, equilibrated in 20% sucrose, sectioned at 20 μ m, and processed for *in situ* hybridization as previously described³⁰. Templates for probes to *trkA*²⁰, *trkB*²¹, and *trkC* each represent a portion of the extracellular domain of the mouse sequences (464, 483, 598 bp, respectively). The mouse *trkC* probe corresponds to residues 635–1,223 of the porcine *trkC* sequence¹⁵ and displays less than 60% nucleotide identity with the homologous region of mouse *trkB*. The probes displayed a comparable specific activity. Motor neurons of the L4 and L5 segments of adult female spinal cord were labelled by application of fluorogold (FG) to cut sciatic nerve 48 h before killing and perfusion of the animals as previously described³¹.

FIG. 4 Messenger RNA for the neurotrophins is present in embryonic limb bud and spinal cord. *A*, Analysis of mRNA for the neurotrophins in E15 limb bud and spinal cord. *B*, Expression of BDNF, NT-3, and NT-5 in developing limb bud and spinal cord. *a, e*, Antisense probe to BDNF. *b, f, g*, Antisense probe to NT-3. *c, h*, Antisense probe to NT-5. *d, i*, Sense probe to NT-5 (control). Sections of E15 rat cervical spinal cord demonstrate a low-level hybridization for BDNF, which is most prominent in the ventral cord (*a*). Intense hybridization for NT-3 is seen in the lateral motor column (lmc in *b*), but no clear signal for NT-5 is observed in cord (*c*). Anterior limb bud at the same age displays hybridization for BDNF and NT-3 in muscle (m in *e* and *f*), as well as expression of BDNF, NT-3 and NT-5 in developing skin (arrowheads *e, f* and *h*). For orientation, floor plate (fp), roof plate (rp), and dorsal root ganglion (drg) are noted in the control section of cord (*d*); *b*, bone structure. Scale bars, 100 μ m.

METHODS. For northern blots, 10 μ g poly(A)⁺-selected RNA from rat E15 spinal cord and whole limb bud was electrophoresed into a formaldehyde-1.2% agarose gel, and blotted onto nylon membranes. Membranes were hybridized with ³²P-labelled neurotrophin-specific DNA probes and then treated as described³². *In situ* hybridization was done as described in Fig. 3. Probes to rat BDNF and NT-3 were as previously described³⁰, and the probe for rat NT-5 represents residues 500–795 of the published sequence¹⁰.



by different factors. In the future, BDNF, NT-3 and/or NT-5 may be used to slow or arrest motor neuron death in neurodegenerative diseases of the anterior horn. □

Received 6 November 1992; accepted 11 March 1993.

- Oppenheim, R. W. *Trends Neurosci.* **12**, 252–255 (1989).
- Levi-Montalcini, R. *Science* **237**, 1154–1162 (1987).
- Barde, Y.-A. *Neuron* **2**, 1525–1534 (1989).
- Yan, Q., Elliott, J. & Snider, W. D. *Nature* **360**, 753–755 (1992).
- Oppenheim, R. W., Qin-Wei, Y., Prevett, D. & Yan, Q. *Nature* **360**, 755–757 (1992).
- Sendtner, M., Holtmann, B., Kolbeck, R., Thoenen, H. & Barde, Y.-A. *Nature* **360**, 757–759 (1992).
- Thoenen, H. *Trends Neurosci.* **14**, 165–170 (1991).
- Arakawa, Y., Sendtner, M. & Thoenen, H. *J. Neurosci.* **10**, 3507–3515 (1990).
- Bloch-Gallego, E. *et al. Development* **111**, 221–232 (1991).
- Berkemeier, L. R. *et al. Neuron* **7**, 857–866 (1991).
- Camu, W. & Henderson, C. E. *J. Neurosci. Meth.* **44**, 59–70 (1992).
- Hynes, M., Gitt, M., Barondes, S. H., Jessell, T. & Buck, L. B. *J. Neurosci.* **10**, 1004–1013 (1990).
- Ericson, J., Thor, S., Edlund, T., Jessell, T. M. & Yamada, T. *Science* **256**, 1555–1560 (1992).
- Martinou, J.-C., Martinou, J. & Kato, A. C. *Neuron* **8**, 737–744 (1992).
- Lamballe, F., Klein, R. & Barbacid, M. *Cell* **66**, 967–979 (1991).
- Soppet, D. *et al. Cell* **65**, 895–903 (1991).
- Quinto, S. P. *et al. Cell* **65**, 885–893 (1991).
- Klein, R. S. *et al. Development* **109**, 845–850 (1990).

- Schecterson, L. C. & Bothwell, M. *Neuron* **9**, 449–463 (1992).
- Martin-Zanca, D., Barbacid, M. & Parada, L. *Genes Dev.* **4**, 683–694 (1990).
- Klein, R. *et al. EMBO J.* **8**, 3701–3709 (1989).
- DiStefano, P. S. *et al. Neuron* **8**, 983–993 (1992).
- Maisonpierre, P. C. *et al. Neuron* **5**, 501–509 (1990).
- Ernfors, P. & Persson, H. *Eur. J. Neurosci.* **3**, 953–967 (1991).
- Rassendren, F. A., Bloch-Gallego, E., Tanaka, H. & Henderson, C. E. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7194–7198 (1992).
- Martinou, J. C., Le Van Thai, A., Cassar, G., Roubinet, F. & Weber, M. J. *Devl Brain Res.* **52**, 175–181 (1990).
- Vogel, K. S. & Davies, A. M. *Neuron* **7**, 819–830 (1991).
- Fonnum, F. J. *Neurochem.* **24**, 407–409 (1975).
- Kawasaki, S. E. & Wang, M. K. in *PCR Technology: Principles and Applications for DNA Amplification* (ed. Erlich, H. A.) 89–97 (Stockton, New York, 1989).
- Phillips, H. S. *et al. Science* **250**, 290–294 (1990).
- Baranowski, A. P., Anand, U. & McMahon, S. B. *Neurosci. Lett.* **141**, 53–56 (1992).
- Rosenthal, A. *et al. Neuron* **4**, 767–773 (1990).

ACKNOWLEDGEMENTS. We thank D. Cooper and S. Barondes (anti-L14 serum) and T. Yamada, T. Jessell and T. Edlund (anti-Islet-1) for generous gifts of antibodies, F. Wurm, H. Parshad, R. Rockwell, L. Burton, E. Martin and A. Gorrell for purified recombinant neurotrophins, C. Schmeltzer, D. Shelton and W.-P. Chan for other factors, M. Vasser, P. Ng and P. Jhurani for synthetic oligonucleotides, M. Bartholdi and M. Rehse for cell sorting, L. Parada for trkB and trkC probes, J. Winslow for trkC probe and M. Epstein, M. Ynes and A. Lumsden for helpful suggestions. This work was financed in part by grants from the Centre National de la Recherche Scientifique, the Institut National de la Santé et de la Recherche Médicale, the Association Française contre les Myopathies (AFM) and the Institut pour la Recherche sur la Moelle Epinière (IRME).

Affinity maturation leads to differential expression of multiple copies of a κ light-chain transgene

Francisco Lozano, Cristina Rada, John M. Jarvis & César Milstein

Medical Research Council, Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

TRANSGENIC animals containing rearranged heavy or light chains are used to study the process of hypermutation, which characterizes the maturation of the antibody response¹⁻³. LK6 mice contain five copies of a transgene coding for a light chain produced in response to the hapten 2-phenyloxazalone⁴. We have selected hybridomas from secondary responses that express the transgene as the only light chain². Some of these hybridomas contain transgene copies carrying mutations known to improve antibody affinity⁵. We have analysed the expression of the five transgene copies in those hybridomas. We report here that the somatic hypermutation process can affect the successful expression of antibody light-chain transgenes. When mutations that improve the antibody affinity appear in one transgene copy, antigenic selection favours cells that downregulate the other copies at multiple levels of gene expression, including examples where nonsense mutations correlate with a drop in messenger RNA level.

Out of the seven hybridomas expressing the κ transgene in antibodies from secondary response to 2-phenyloxazalone², we selected four for further study because most of their transgene copies contained several point mutations which often result in changes in charged residues, detectable by isoelectric focusing (IEF) (Fig. 1). The other three hybridomas did not contain mutations or mainly contained mutations not leading to charge changes. The four selected hybridomas are listed in Table 1. Generally, the more intense component does not correspond to the number of transgene copies of the expected IEF mobility (Fig. 1). For instance, the band corresponding to transgene copies with wild-type charge (three out of four copies) is very weak in hybridoma NQT17/29 (Fig. 1a). Similar comments can be made about the patterns of hybridomas NQT17/45 and NQT17/30. Of particular interest are hybridomas NQT17/10 and NQT17/29. Both carry a copy with predicted acidic charge (A and B, respectively) and containing His 34 $\rightarrow$ Asn and

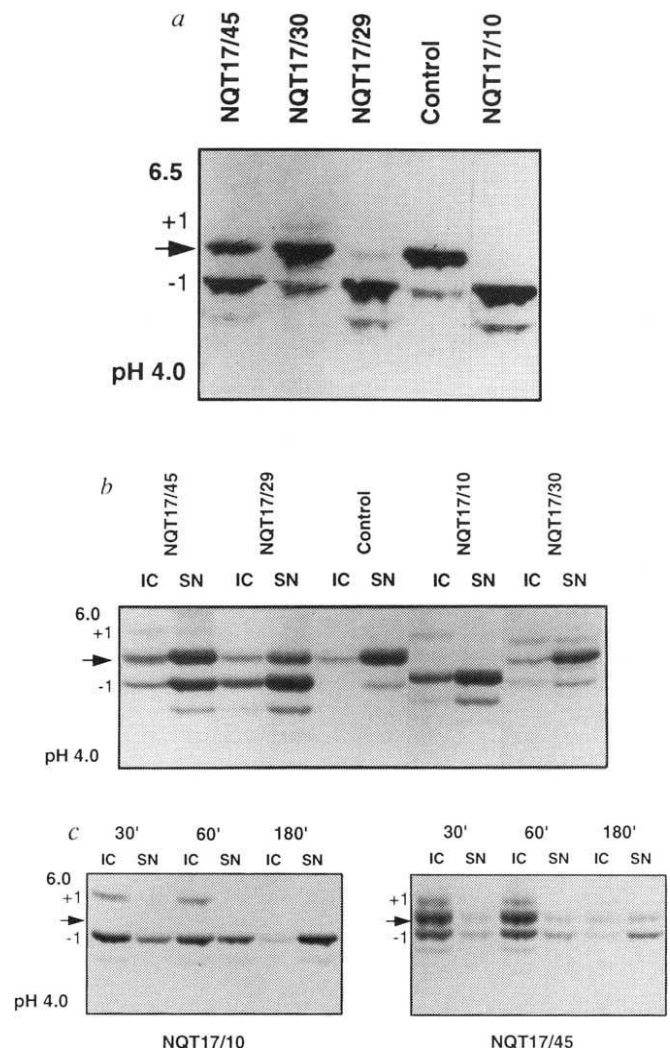


FIG. 1 Isoelectric focusing analysis of transgenic κ light chains synthesized by Lk6 hybridomas. The Lk6 mouse line contains transgenes made of a V κ Ox1-J κ 5 mouse V-segment and a rat C κ . The derived hybridomas are specific for the hapten 2-phenyloxazalone. On the left, the mobility of the wild-type transgenic κ light chains is indicated by an arrow. The mobility of basic (+1) and acidic (-1) forms of the transgenic κ light chains is also indicated. Control refers to the unmutated transgenic κ light chains secreted by the NQT15/25 hybridoma². The minor more negatively charged component associated with the IEF bands results from deamidation. a, Transgenic κ light chains associated to endogenous heavy chains in secreted antibodies. Hybridoma culture supernatants were incubated with protein A-Sepharose CL-4B beads. The immunoprecipitated material was eluted under reducing conditions and resolved by isoelectrofocusing (IEF) on agarose-urea gels. Proteins were transferred to nitrocellulose membranes and developed using a biotinylated anti-rat κ monoclonal antibody (MRC OX-12, Serotec) followed by incubation with peroxidase-labelled avidin. b, Autoradiograms of total ³⁵S-labelled transgenic κ chain detected intra- and extracellularly. Culture supernatants (SN) and detergent-soluble cell lysates (IC) from ³⁵S-methionine-labelled hybridomas were incubated with MRC OX-12 coupled to Sepharose beads. The immunoprecipitated material was eluted under reducing conditions and subjected to IEF on agarose-urea gels. c, Intracellular degradation of transgenic κ light chains in NQT17/10 and NQT17/45 hybridomas. Cells were pulsed for 30 min with ³⁵S-methionine and chased

at 30, 60 and 180 min. At each time, culture supernatants (SN) and detergent-soluble cell lysates were immunoprecipitated and analysed as in b. METHODS. For metabolic labelling, 4×10^6 cells were washed twice and resuspended in 1.5 ml methionine-free Dulbecco's modified Eagles medium (DMEM) supplemented with 5% dialysed fetal calf serum. After 30 min incubation at 37 °C, 200 μ Ci ³⁵S-methionine (Amersham) were added, and cells were cultured for 3 h at 37 °C. In pulse-chase experiments, cells were incubated with the same amount of ³⁵S-methionine for 30 min, washed twice and resuspended in 1.5 ml fresh complete DMEM. Culture supernatants were collected, and cell pellets washed twice in ice-cold PBS before resuspending for 30 min in 1 ml ice-cold lysis buffer containing 1% Nonidet P-40 and protease inhibitors. Detergent-soluble lysates or culture supernatants were precleared with BSA-coupled Sepharose beads and immunoprecipitated with 50 μ l of packed protein A-Sepharose CL-4B or MRC OX-12 coupled to Sepharose beads. BSA and MRC OX-12 were coupled to Sepharose CNBr-activated beads following the manufacturer's instructions (Pharmacia, Uppsala). After 1-2-h incubation at 4 °C under gentle orbital rotation, immunoprecipitates were sequentially washed in lysis buffer and TNE buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM EDTA). Immunoprecipitated material was eluted at room temperature in 9.8 M urea, 10 mM dithiothreitol, 50 mM Tris-HCl, pH 8.0. Sample eluates were analysed by IEF in 1.8% w/v agarose gels containing 6 M urea and 6% v/v of a mixture of Pharmalyte carrier ampholytes (1.8% pH 3-10, and 4.2% pH 4-6.5) (Pharmacia, Uppsala), as described¹⁷. Gels were fixed, dried and autoradiographed. Unlabelled transgenic light chains were transferred from agarose-urea gels to nitrocellulose membranes (Schleicher & Schuell, Dassel, Germany) by capillary blotting. The membrane was blocked for 30 min at room temperature with 5% BSA and 3% fat-free dry milk powder (in PBS) and then incubated for 1 h with a 1:200 dilution of biotinylated MRC OX-12. After three washes with PBS, 0.25% gelatine, the membrane was incubated for 1 h with a 1:1,000 dilution peroxidase-conjugated avidin (Sigma). After three PBS, gelatine washes, a diaminobenzidine NiCl₂ substrate solution was added to develop antigen-bound antibody¹⁸.

Tyr 36 → Phe mutations, which increase anti-oxazolone binding affinity by a factor of 10 (ref. 5). In each, the acidic product (-1) is by far the most prominent in the culture supernatants (Fig. 1a). In hybridomas NQT17/10 and NQT17/45, expected basic bands (corresponding to copy B and D, respectively) are not detectable in culture supernatants (Fig. 1a). The missing bands are nonetheless more prominent in the intracellular compartment (Fig. 1b), suggesting intracellular degradation. This was demonstrated by pulse-chase experiments (Fig. 1c).

To assess the contribution of each transgene copy to the mRNA pool, we sequenced 20 random cDNA clones from each hybridoma (Table 2). There seems to be an uneven representation of the different copies. The extreme cases are NQT17/10 and NQT17/29, where three and one of the copies, respectively, are not represented in the complementary DNA pool. In contrast, the copies with advantageous mutations are the most prominently represented. After sequencing 10 clones derived by direct amplification of genomic DNA, all five expected transgene copies are found at similar frequencies. Among the copies that did not yield cDNA clones, there were three that included nonsense mutations (Fig. 2c). In two hybridomas, all transgene copies are somatically mutated in spite of the fact that they occur in opposite orientations (Fig. 2a). This does not support the suggestion that hypermutation is linked to the direction of replication⁶.

The imbalance in mRNA expression cannot be attributed to gross deletions or rearrangements. Southern blot analysis with *Bam*HI restriction digest, and probes specific for the J-C intron and 3' enhancer regions reveal no differences between one hybridoma expressing all five copies and NQT17/10, which expresses only two (Fig. 2b). Extensive sequence analysis in the upstream segments of the unexpressed copies of NQT17/10 reveals only a few point mutations, but none in segments invol-

TABLE 2 Genome copies and mRNA species of the Lk transgene

Hybridoma	Variant transgene	Relative number of PCR clones sequenced	
		cDNA	DNA
NQT17/10	A	14/20	4/25
	B	6/20	4/25
	C*	0/20	5/25
	D*	0/20	9/25
	E*	0/20	3/25
NQT17/29	A	2/19	2/10
	B	8/19	3/10
	C	7/19	2/10
	D	2/19	4/10
	E†	0/19	1/10
NQT17/30	A	5/19	ND
	B	3/19	
	C	5/19	
	D	2/19	
	E†	2/19	
NQT17/45	A	4/19	ND
	B	5/19	
	C	6/19	
	D	2/19	
	E	2/19	

Cloning and sequencing of cDNA was as described¹⁵. Genomic DNA was isolated from hybridomas using standard procedures¹⁶ and subjected to PCR amplification. The primers used were VKOXBACK (5'-CGGGGAATTCT-CAGCTTCCTGCTAATCAG-3'), which primes the leader of V α 1 gene, and ELKBAM (5'-CGCGGATCCCTTTCTATCCTGAAGTTCCT-3'), which primes the J-C intron segment. PCR products were digested with *Eco*RI and *Bam*HI, purified, cloned into M13mp18, and the V segment sequenced with Sequenase kit using the JK5FOR primer. ND, not done.

* Includes a nonsense mutation.

† Unmutated V segment.

TABLE 1 Amino-acid substitutions and relative charge of transgenic κ light chains

Hybridoma	Transcribed transgenes	AA substitutions	Charge change
NQT17/10	A	34H → N , 36Y → F , 95P → S	~-1
	B	101G → R	+1
NQT17/29	A	31S → T, 32Y → N, 77S → N, 96L → V	0
	B	33M → I, 34H → N , 36Y → F	~-1
	C	60A → V	0
	D	60A → T, 74T → S	0
NQT17/30	A	32Y → N, 36Y → D, 48I → T, 82D → E, 86Y → N	-1
	B	36Y → N, 41G → D, 53K → R	-1
	C	Silent mutations	0
	D	13A → V, 26S → N, 47M → R	+1*
	E	Unmutated	0
NQT17/45	A	62F → L, 83A → V, 98F → V	0
	B	26S → N, 35M → L, 36Y → S, 48I → T	0*
	C	60A → D	-1
	D	31S → R, 77S → N, 87Y → C	+1
	E	33M → I	0

The sequences of the transcribed copies of the Lk transgene for all the hybridomas were updated from cDNA clones generated by polymerase chain reactions (PCR) amplification of specifically primed first-strand cDNA (see Table 2 for details). Only amino-acid substitutions introduced by coding point mutations in the V region of original Lk transgenes are shown. Concomitant silent mutations present in each transgene copy are as described². The transgene copy D from hybridoma NQT17/29 has not been previously reported. This copy contains no silent mutations. The presence of amino-acid substitutions characteristic of affinity maturation in the anti-phOx immune response (34H → N and 36Y → F) is highlighted (bold). In some cases (*) the predicted charge change may be modified by the presence of a putative N-glycosylation site. But no significant change in the IEF patterns in Fig. 1 were detected after neuraminidase treatment.

ved in transcription (for example, TATA box and octamer); (Fig. 2c). In contrast, the -80 mutation in copy E, which may distort the splicing pattern by generating a new acceptor site, and the mutation in the penultimate amino acid of the leader peptide of the same copy E, leading to early termination, may both have affected the mRNA stability⁷.

Thus, the greatest imbalance in the expression of the different transgene copies is observed when one copy acquires mutations that improve antigen binding. This renders the other copies either dispensable, mere passengers, or even disadvantageous, because they would compete for heavy-chain binding. The imbalance is best detected at the protein level, where both downregulation of specific mRNAs and accumulation of deleterious mutations that impair the stability or secretion of light chain combine their effects. Effectively, the biological selection during maturation seems to favour a gradual exclusion phenomenon which, as in isotypic, allelic and idiotypic exclusion, favours the expression of a single copy of multiple alternatives. The point has relevance regarding a recent controversy on the role of cellular selection in allelic exclusion⁸⁻¹¹. Although we do not suggest that the mechanisms of exclusion *per se* rely primarily on this type of event, our results strongly imply that cellular selection during development and/or maturation may play an important role, at least in the elimination of leaky events.

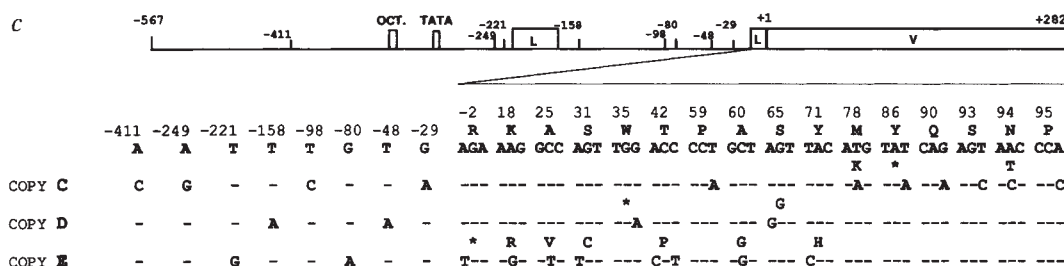
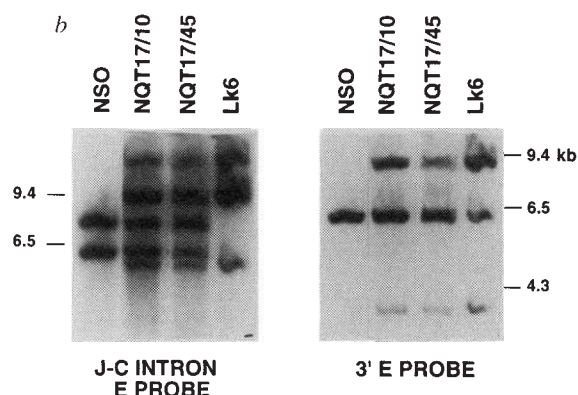
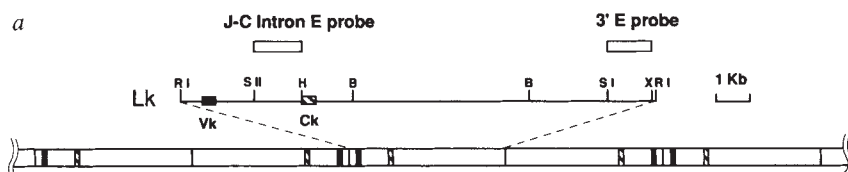
Downregulation of transgenes has previously been observed during development from neonates to adult animals¹². This could easily have been explained by general downregulation of the whole transgene complex in a concerted way. Here, downregulation of the mRNA is selective within a set of almost identical genes. Why does the mRNA pool not result from an equal contribution of each transgene copy? We have no evidence on whether this is regulated at the level of transcription or post-transcriptionally. We have not been able to identify positively any mutations in the promoter region that could explain

FIG. 2 *a*, Diagrammatic map of the Lk transgene and configuration of the five integrated transgene copies in Lk6 mice. Restriction sites are abbreviated: H, *Hpa*I; RI, *Eco*RI; SI, *Sac*I; SII, *Sac*II; X, *Xho*I.

b, Southern blot analysis of Lk6 hybridomas. Genomic DNA was digested with *Bam*HI and probed for the J-C intron enhancer region, as well as for the 3' enhancer region. Genomic DNA samples from the parental myeloma cell line (NSO) and Lk6 spleen (Lk6) are also included.

c, Point mutations found in the untranscribed copies of hybridoma NQT17/10. Includes sequencing data from the nucleotide -567 to +282 counting from the first base of the first residue of the mature VkOx1 light chain. Nucleotides that differ from the original transgene are shown. Dashes and gaps indicate identity to the reference sequence. Codons are numbered according to ref. 19. The unmutated nucleotide sequence is that of the H3 gene²⁰.

METHODS. For Southern analysis, genomic DNA was digested with *Bam*HI (New England Biolabs), size-fractionated in 0.8% agarose gel and transferred onto Hybond Plus membranes (Amersham) by capillary alkaline blotting. Membranes were prehybridized for 30 min at 65 °C with 7% SDS in 0.5 M sodium phosphate buffer pH 7.2, 2 mM EDTA. The DNA probes used (a *Sac*II-*Hpa*I fragment of the mouse J-C intron enhancer region, and a *Sac*I-*Xba*I fragment of the 3' enhancer region) were radiolabelled with [³²P]dCTP (Oligolabelling kit, Pharmacia) and hybridized overnight at 65 °C in the above-mentioned solution. Membranes were washed with 1% SDS in 0.5 M sodium phosphate buffer, pH 7.2. For cloning and further sequencing, genomic DNA was isolated from NQT17/10 using standard procedures¹⁶ and subjected to PCR amplification. The primers used were 5PRIMAOX BACK (5'-CCCGA-ATTCTAGTCAGTACTCTC-3'), which primes near the 5' end of the Lk transgene, and JK5FORBAM (5'-CGTTTCGGATCCAGCTTGGTCCCAG-3'), which primes the Jκ5 segment. PCR products were digested with *Eco*RI and *Bam*HI, purified, cloned into M13mp18 and sequenced with Sequenase kit (USB) using the JK5FOR (5'-CGTTAGATCTCCAGCTTGGTCCCAG-3') VKOX30 (5'-GATTGCTGGAGACTGGG-3') and the LPOX (5'-ACTGATTAGCAGGAAGC-3') primers.



such downregulation. Is it possible that the mutations that we detect downregulate mRNA expression or stability? Nonsense mutations within exons can reduce the level of some mRNAs^{13,14}. The stop codons found in NQT17/10, copy C, D and E, are consistent with the suggestion that there may be a direct coupling between translation and the mRNA levels⁷. There also remains the possibility that epigenetic modifications could downregulate those genes that are either dead weight or disadvantageous for the expression of antibody. None of these possibilities is exclusive of others, and different transgene copies

may be downregulated by alternative routes. What is clear is that the study of the downregulation of individual transgene copies (a very frequent event in our studies) could shed considerable light on alternative mechanisms of quantitative changes in gene expression. Whichever is the mechanism, we find it intriguing that antigen can recognize such an advantage. This suggests that antigenic selection not only operates in terms of the increased affinity of a particular molecule, but also in terms of the number of antibody receptors that the cell expresses. □

Received 17 December 1992; accepted 16 March 1993.

- Hackett, J., Rogerson, B. J., O'Brien, R. L. & Storb, U. *J. exp. Med.* **172**, 131-137 (1990).
- Sharpe, M. J., Milstein, C., Jarvis, J. M. & Neuberger, M. S. *EMBO J.* **10**, 2139-2145 (1991).
- Sohn, J., Gerstein, R. M., Hsieh, C., Lerner, M. & Selsing, E. *J. exp. Med.* **177**, 493-504 (1993).
- Meyer, K. B., Sharpe, M. J., Surani, M. A. & Neuberger, M. S. *Nucleic Acids Res.* **18**, 5609-5615 (1990).
- Berek, C. & Milstein, C. *Immun. Rev.* **96**, 23-41 (1987).
- Rogerson, B., Hackett, J., Peters, A., Haasch, D. Storb, U. *EMBO J.* **10**, 4331-4341 (1991).
- Goldfarb, D. & Michaud, N. *Trends Cell Biol.* **1**, 20-24 (1991).
- Kitamura, D. & Rajewsky, K. *Nature* **356**, 154-156 (1992).
- Wabl, M. & Steinberg, C. *Nature* **359**, 370-371 (1992).
- Oancea, A. & Shulman, M. J. *Nature* **359**, 371 (1992).
- Rajewsky, K., Ehlich, A., Müller, W., Taki, S. & Kitamura, D. *Nature* **359**, 171-172 (1992).
- Pettersson, S., Sharpe, M. J., Gilmore, D. R., Surani, M. A. & Neuberger, M. S. *Int. Immun.* **1**, 509-516 (1989).
- Cheng, J., Fogel-Petrovic, M. & Maquat, L. E. *Molec. cell. Biol.* **10**, 5215-5225 (1990).

- Urbach, B., Mitchell, P. J., Ciudad, C. J. & Chasin, L. A. *Molec. cell. Biol.* **9**, 2868-2880 (1989).
- Sharpe, M. J., Neuberger, M. S., Pannell, R., Surani, M. A. & Milstein, C. *Eur. J. Immun.* **20**, 1379-1385 (1990).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
- Nespolo, A. *et al. Electrophoresis* **10**, 34-40 (1989).
- Harlow, E. & Lane, D. A. in *Antibodies: A Laboratory Manual* 509 (Cold Spring Harbor Laboratory Press, New York, 1988).
- Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. in *Sequences of Proteins of Immunological Interest* (US Government Printing Office, Bethesda, 1991).
- Milstein, C., Even, J., Jarvis, J. M., Gonzalez-Fernandez, A. & Gherardi, E. *Eur. J. Immun.* **22**, 1627-1634 (1992).

ACKNOWLEDGEMENTS. We thank M. Sharpe and M. Neuberger for suggestions and materials, and R. Pannell for tissue culture assistance. F.L. is the recipient of a Celltech Fellowship and C.R. is supported by the Spanish Ministry of Science and Education.

Substitution of three amino acids switches receptor specificity of $G_q\alpha$ to that of $G_i\alpha$

Bruce R. Conklin*, Zvi Farfel†, Kevin D. Lustig*, David Julius* & Henry R. Bourne*‡

* Departments of Pharmacology and Medicine, Program in Cell Biology, and the Cardiovascular Research Institute, S1212 Box 0450, University of California, San Francisco, California 94143, USA

† Department of Medicine E, Chaim Sheba Medical Center, Tel Hashomer, Tel Aviv University, Tel Aviv 52621, Israel

AGONIST-BOUND receptors activate heterotrimeric ($\alpha\beta\gamma$) G proteins by catalysing replacement of GDP bound to the α -subunit by GTP¹⁻⁵. Mutations in the C terminus of the α -subunit^{6,7}, its covalent modification by pertussis toxin-catalysed ribosylation of ADP⁸, peptide-specific antibodies directed against it⁹⁻¹¹, and peptides mimicking C-terminal sequences¹², all inhibit receptor-mediated activation of G proteins. The logical prediction—that specific amino-acid residues at the C-termini of α -subunits can determine the abilities of individual G proteins to discriminate among specific subsets of receptors—has so far not been tested experimentally. Different hormone receptors specifically activate G_q or G_i , whose α -subunits (α_q or α_i) stimulate phosphatidylinositol-specific phospholipase C or inhibit adenylyl cyclase, respectively¹⁻⁵. Here we replace C-terminal amino acids of α_q with the corresponding residues of α_{i2} to create α_q/α_{i2} chimaeras that can mediate stimulation of phospholipase C by receptors otherwise coupled exclusively to G_i . A minimum of three α_{i2} amino acids, including a glycine three residues from the C terminus, suffices to switch the receptor specificity of the α_q/α_{i2} chimaeras. We propose that a C-terminal turn, centred on this glycine, plays an important part in specifying receptor interactions of G proteins in the $G_i/G_o/G_z$ family.

We chose α_q and α_{i2} as candidates for transference of receptor specificity because G_q and G_i discriminate among different subsets of receptors and because they regulate distinct and easily measurable second messenger pathways. The first set of chimaeras tested were α_q subunits in which C-terminal residues were progressively replaced with corresponding residues from α_{i2} ; each chimaera was designated by the number of α_{i2} residues that replace corresponding residues in α_q (Fig. 1). Among the 24 C-terminal residues of α_q and α_{i2} , there are 12 non-identical amino acids. In the series from qi1 to qi23, each chimaera differs from the preceding one by substitution of an amino acid at a single position, with the exception of qi13 and qi18, which differ at two positions.

In human embryonic kidney (HEK)-293 cells, we transiently co-expressed DNA encoding each α_q/α_i chimaera with DNAs encoding a panel of three different receptors, each of which couples predominantly or exclusively to G_i . Two of these, the A_1 adenosine (A_1R) and D_2 dopamine (D_2R) receptors, activate G_i in HEK-293 cells¹³ and do not affect phospholipase C (PLC) even when α_q is overexpressed (see qWT in Fig. 1b). The third, the α_2 -adrenoceptor (α_2-AR), shares with a small but growing class of receptors¹⁴⁻¹⁷ the ability to activate more than one G protein; it preferentially activates G_i but can also activate G_q when α_q is overexpressed in HEK-293 cells¹⁸ (see qWT in Fig. 1b). Therefore α_2-AR stimulation of PLC indicates that the α_q/α_i chimaera is functionally expressed, whereas A_1R or D_2R stimulation of PLC indicates that the chimaera has gained an ability to interact with receptors that couple exclusively to G_i .

Measurements of receptor-stimulated PLC activity in cells expressing qi1–qi23 showed that replacing just three of the last four amino acids of α_q with the corresponding α_{i2} residues

suffices to alter receptor specificity (see qi4 in Fig. 1). More extensive substitutions did not further increase PLC stimulation. Indeed, substitution of α_{i2} residues in excess of those present in qi9 reduced the extent of PLC stimulation by the D_2R and A_1R , although these chimaeras still allowed the α_2-AR to stimulate PLC.

Expression of chimaeras qi4–qi11 raised basal PLC activity (Fig. 1), suggesting that these α -subunits are constitutively active. Although reproducible, the activation is weak, elevating basal PLC only twofold as opposed to a sixfold increase caused by α_q with a GTPase-inhibiting mutation measured under identical conditions¹⁸. Mutations in the C termini of α_o and α_s produce similarly weak constitutive activities, probably by accelerating GDP dissociation^{19,20}.

Although indistinguishable in their abilities to mediate inhibition of adenylyl cyclase through endogenous G_i proteins¹³, the

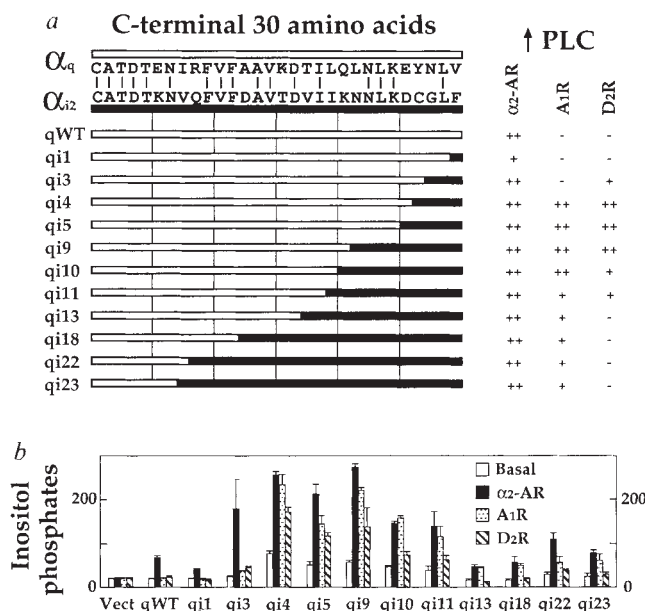


FIG. 1 G_q -coupled receptors stimulate PLC in cells transfected with α_q/α_i chimaeras. **a**, Amino-acid composition of the chimaeras, with a summary of PLC stimulation by the α_2-AR , A_1R or D_2R in three or more experiments. In bars representing the C-terminal 30 residues, unfilled segments designate α_q residues and cross-hatched segments designate α_{i2} residues. A minus sign denotes no significant stimulation of PLC; + denotes a 1–2-fold stimulation; and ++ denotes a 2–5-fold stimulation. **b**, A representative experiment showing the effect of agonists of G_i -coupled receptors on inositol phosphate (InsP) formation in HEK-293 cells transfected with pcDNA (vector), qWT or individual α_q/α_{i2} chimaeras. Agonists: 10 μM UK-14304 for α_2-AR (filled bars); 10 μM (+)-N6-(2-phenyl-isopropyl)adenosine (PIA) for A_1R (stippled bars); and 10 μM quinpirole for D_2R (cross-hatched bars). **METHODS.** Chimaeras of murine α_q (ref. 32) and murine α_{i2} (ref. 33) cDNAs in pcDNA (Invitrogen) were produced by polymerase chain reactions (PCR) using either a 3-primer fusion strategy³⁴ (for qi13–qi23) or by including the entire added sequence and *Nsi*I site in the 3' PCR primer (for qi1–qi11). The PCR product was digested with *Eco*RI and *Nsi*I, resulting in 178-bp fragment which was subcloned into α_q -pcDNA, and confirmed by DNA sequencing. HEK-293 cells (American Type Culture Collection; CRL-1537) were transiently transfected by the DEAE-dextran method as described¹⁸ in 60-mm dishes (10^6 cells per dish) with the indicated cDNAs: 1 μg α_2-AR -pcDNA (ref. 35); 1 μg D_2R -pcDNA-I (ref. 36); 1 μg A_1R -CDM7 (ref. 37) (provided by P. D. Garcia); 3 μg qWT or 3 μg α_q/α_{i2} chimaera. Each 60-mm dish was split into 12 wells of a 24-well plate 24 h after transfection [3H]inositol (2 μCi per well) was added to the growth medium 48 h after transfection. After a 24-h labelling period, cells were stimulated with receptor agonists and [3H]IPs were resolved on Dowex columns as described¹⁸. Results are expressed as c.p.m. [3H]IP ($\times 1,000$) divided by the sum of the c.p.m. in both the [3H]InsP and [3H]inositol fractions. Data represent the mean $\pm$ s.d. or triplicate determinations in a single experiment; 3–7 additional experiments gave similar results.

‡ To whom correspondence should be addressed

A₁R and the D₂R differ significantly in their ability to stimulate chimaeras q13 and q10–q23 (Fig. 1). Concentration-response curves for PLC stimulation by agonists specific for these receptors dramatically illustrate these differences (Fig. 2). Thus wild-type- α_q (qWT) mediated PLC stimulation only by an α_2 -AR agonist, UK-14301, whereas q13 mediated responsiveness to UK-14301 and also to the A₁R agonist, (+)-N6-(2-phenylisopropyl)adenosine (PIA), but not to the D₂R agonist, quinpirole. Chimaera q15 mediated PLC stimulation by all three receptor-specific agonists, and q13 mediated a response to the α_2 -AR agonist, a weaker response to the D₂R agonist, and no significant response to the A₁R agonist. These patterns point to potential differences, undetectable in other assays, among the G-protein-binding sites of these G_i-coupled receptors.

Chimaeras in which the last five residues of α_q were replaced with the corresponding amino acids of α_z and α_o (Fig. 3a) allowed PLC responses to A₁R and D₂R agonists comparable to those mediated by q15 (Fig. 3b). A chimaera in which the last five residues of α_q were replaced with the corresponding amino acids of α_s did not allow G_s-coupled receptors to stimulate PLC (data not shown). Because α_z and α_o , unlike α_s , belong to the α_i protein family² and respond to similar subsets of receptors^{3,13}, these results are not surprising.

The glycine at the -3 position is important for ADP-ribosylation of α_o by pertussis toxin (PTx)²¹. PTx-catalysed attachment of ADP-ribose to a cysteine residue at the -4 position⁸ prevents proteins in the G_i family from interacting with receptors. Thus it is not surprising that PTx inhibits stimulation of PLC activity by the α_2 -AR and D₂R in cells expressing chimaera q14, which contains a cysteine and a glycine at the -4 and -3 positions, respectively. PTx fails to prevent D₂R-mediated activation of PLC in cells expressing qz5, a chimaera in which the toxin's target residue is replaced by isoleucine (Fig. 3c).

In combination with results from other laboratories^{22,23}, our observations support a model in which the distinctive three-

dimensional shape of a short C-terminal region of G α plays a critical role in receptor-G protein interactions. This region in α_i proteins probably forms a β -turn. NMR analysis of an undecapeptide representing the C terminus of the α -subunit (α_i) of retinal G_i, described in the accompanying paper²², revealed a β -turn centred on the glycine residue that is conserved at the -3 position in α -subunits of the entire α_i/α_o family; in our experiments, this glycine is common to all the chimaeras that switched receptor specificity (Fig. 3a).

a	C-terminal amino acids				
	-5	-4	-3	-2	-1
α_q	E	Y	N	L	V
α_{i2}	D	C	G	L	F
α_o	G	C	G	L	Y
α_z	Y	I	G	L	C
α_s	Q	Y	E	L	L

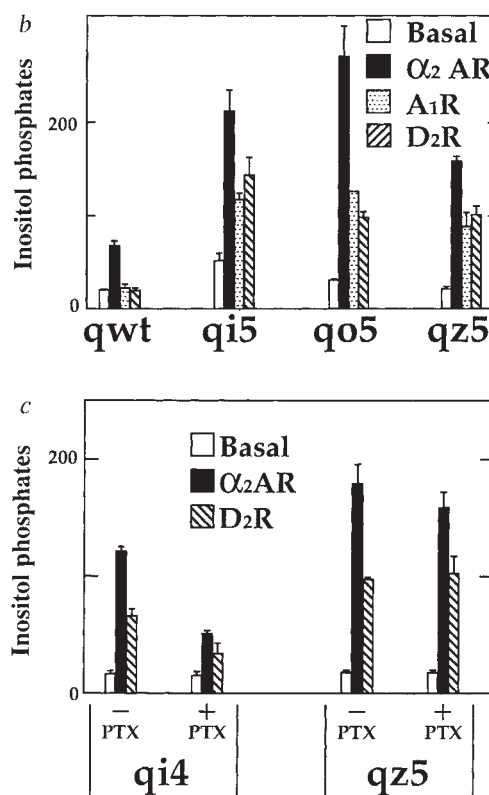


FIG. 3 Stimulation of PLC by chimaeras created by replacing the five C-terminal amino acids of α_q with the corresponding residues of α_{i2} , α_o or α_z . a, Comparison of the five C-terminal amino acids of α_q , α_{i2} , α_o , α_z and α_s . b, Stimulation of InsP formation by agonists of transfected G_i-coupled receptors in HEK-293 cells expressing qWT, q15, qo5, and qz5. c, Effect of PTx on InsP formation by the α_2 -AR or D₂R in HEK-293 cells expressing q14 or qz5. PTx (200 ng ml⁻¹) was added to the cell growth medium 18 h before the experiment. Agonists in b and c were the α_2 -AR agonist UK-14304, 10 μ M (filled bars); the A₁R agonist PIA, 10 μ M (stippled bars); or the D₂R agonist quinpirole, 10 μ M (cross-hatched bars). Immunoblot analysis of cells transfected under identical conditions showed that qWT is expressed at a level 50% higher than q15, qo5 or qz5 (data not shown), indicating that the new functions of the chimaeras are not due to increased expression. Transfections, InsP assays and chimaera syntheses are described in the legend to Fig. 1. Data are expressed as the mean $\pm$ s.d. of triplicate determinations in a single experiment; three additional experiments gave similar results.

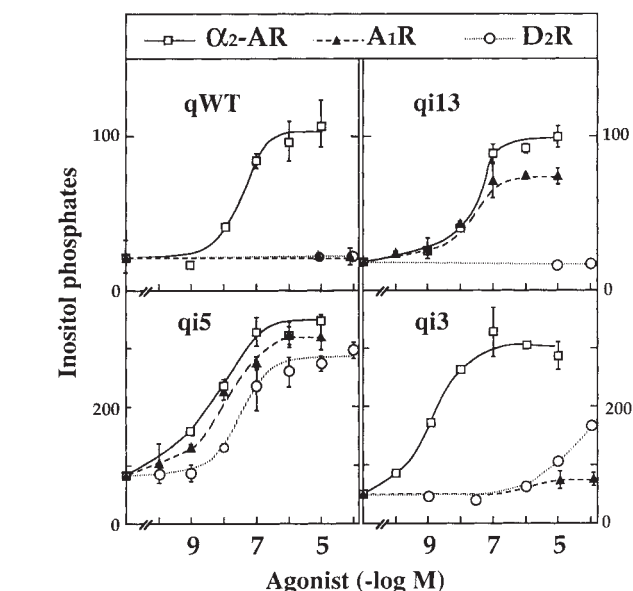


FIG. 2 Concentration-response curves for stimulation of PLC by three G_i-coupled receptors in HEK-293 cells transfected with qWT, q13, q15 or q13. Cells were treated with the indicated concentration of the α_2 -AR agonist, UK-14304 (squares), the A₁R agonist, PIA (filled triangles), or the D₂R agonist, quinpirole (circles). Immunoblot analysis of cells transfected under identical conditions showed that qWT is expressed at a level ~50% higher than q13, q15 or q13 (data not shown), indicating that the new functions of these chimaeras are not due to enhanced expression of the protein itself. Transfections and InsP assays were done as described in Fig. 1 legend, except that data are expressed as the mean $\pm$ range of duplicate determinations in a single experiment. One to three additional experiments gave similar results.

In addition to its role as a determinant of specificity, the C termini of α -subunits may also serve as an essential trigger for receptor activation of G proteins. This provocative notion is based on evidence that receptors and this region of α -chains undergo coordinated conformational changes during the ligand-binding reaction. Binding of agonist to the α_2 -AR renders the C terminus of the associated α_i inaccessible to antibody²⁴, suggesting that the activated form of the receptor may shield or envelop the C terminus of α_i . NMR analysis, presented in the accompanying paper²², shows that binding of rhodopsin to a C-terminal decapeptide derived from α_i causes a conformational change in the peptide^{22,23}; in the other direction, the same peptide induces in rhodopsin a spectral change like that induced by binding G_i itself²³. Furthermore, mutations of the C terminus cause an apparent increase in the rate of GDP dissociation^{19,20}, indicating that a conformational change in this region may regulate GDP release, the primary event in G-protein activation.

Although the extreme C terminus of the α -subunit plays a significant part in receptor specificity, it is certainly not the sole determinant. Two splice variants of α_o couple to different receptors²⁵, despite the fact their eight C-terminal amino acids are identical^{2,3}. Similarly, the last eight amino acids of α_{i2} and α_i are identical, but the α_2 -AR activates α_{i2} and not α_i (ref. 26). Moreover, G protein $\beta\gamma$ subunits help anchor α -subunits to receptors²⁷⁻²⁹ and differences among β -subunits contribute to specificity of receptor coupling³⁰. Thus specific interactions with receptors also depend upon other regions of α -subunits, probably including but not limited to the N-terminal region that regulates association with $\beta\gamma$ ³¹.

In summary, we have defined a small region of α_{i2} , α_o and α_z that confers upon α_q a new receptor specificity and (in the case of α_{i2} sequence) sensitivity to PTx. We predict that once the other receptor specificity regions of G proteins are identified, G proteins can be designed to harness any G-protein-coupled receptor to the signalling pathway of choice. □

NMR structure of a receptor-bound G-protein peptide

Edward A. Dratz*, Julie E. Furstenau*, Christophe G. Lambert*†, Dennis L. Thireault*, Helen Rarick‡, Theresa Schepers‡, Sergei Pakhlevaniants‡ & Heidi E. Hamm‡§||

* Department of Chemistry and Biochemistry, Montana State University, Bozeman, Montana 59717, USA

Departments of ‡ Physiology and Biophysics and § Ophthalmology, University of Illinois College of Medicine at Chicago, Chicago, Illinois 60680, USA

|| Istituto di Fisiologia Generale e Chimica Biologica, Università di Sassari, 07100 Sassari, Italy

HETEROTRIMERIC GTP-binding proteins (G proteins) regulate cellular activity by coupling to hormone or sensory receptors. Stimulated receptors catalyse the release of GDP from G protein α -subunits¹⁻⁴ and GTP bound to the empty α -subunits provides signals that control effectors such as adenylyl cyclases, phosphodiesterases, phospholipases and ion channels⁴. Three cytoplasmic loops of the activated receptor are thought to interact with three sites on the heterotrimeric G protein to provide high-affinity interaction and catalyse G-protein activation⁵⁻⁸. The carboxyl terminus of the α -subunit is particularly important for interaction with the receptor⁹⁻¹⁴. Here we study the structure of part of the active interface between the photon receptor rhodopsin and the G protein transducin, or G_t , using nuclear magnetic resonance. An 11-amino-acid peptide from the C terminus of the α -subunit of G_t (α_t (340-350)) binds to rhodopsin and mimics the G protein in stabilizing its active form, metarhodopsin II. The peptide α_t (340-350) binds to both excited and unexcited rhodopsin and conformational differences between the two bound forms suggest a mechanism for activation of G proteins by agonist-stimulated receptors. Insight into receptor-catalysed GDP release will have broad application because the GTP/GDP exchange and the intrinsic GTPase activity of GTP-binding proteins constitute a widespread regulatory mechanism¹⁵.

Understanding how G proteins are activated by receptors will be facilitated by high-resolution structural information. Transferred nuclear Overhauser effect (NOE) NMR spectroscopy (Tr-NOESY) was used to study the three-dimensional structure of α_t (340-350) bound to rhodopsin and metarhodopsin II. The Tr-NOESY experiment can provide information on the structure of small bound molecules if they exchange rapidly on and off binding sites on large receptors^{16,17}. The synthetic peptide α_t (340-350) binds to rhodopsin and mimics the action of G_t to stabilize metarhodopsin II (MII; ref. 8) (Fig. 1b). Analogous peptides, α_t (340-350)K341R (with Arg substituted for Lys at 341), or an N-acylated derivative, have a similar effect and were used for Tr-NOESY because of certain advantages (Fig. 1b legend). Tr-NOESY spectra were measured on the peptide in rapid exchange with membrane-bound rhodopsin. Figure 2 shows Tr-NOESY data for α_t (340-350)K341R in the presence of dark rhodopsin in the NH-NH region (Fig. 2a) and in the NH- α H and NH- β H regions (Fig. 2b) of the two-dimensional NMR spectra.

The NOEs provide information on distances between various protons in the bound peptide. Distance geometry, NOE-constrained molecular dynamics, simulated annealing and grid searches yielded a family of very similar structures that were consistent with the NOE distance constraints. Figure 3a shows superpositions of deduced structures of the α_t (340-350)K341R peptide when bound to dark rhodopsin. Figure 3c shows a stereo view of a representative structure which gives the best agreement

Received 9 November 1992; accepted 16 February 1993.

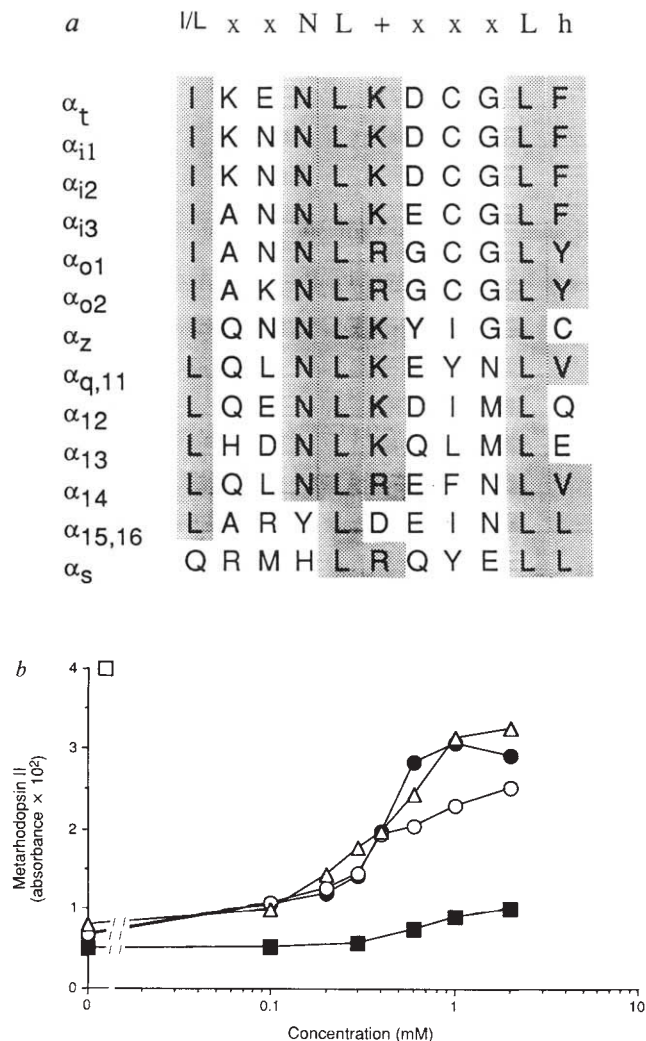
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125-132 (1990).
- Simon, M. I., Strathmann, M. P. & Gautam, N. *Science* **252**, 802-808 (1991).
- Birnbaumer, L. A. *Rev. Pharmac. Toxicol.* **30**, 675-705 (1990).
- Kaziro, Y., Itoh, H., Kozasa, T., Nakafuku, M. & Satoh, T. A. *Rev. Biochem.* **60**, 349-400 (1991).
- Gilman, A. G. A. *Rev. Biochem.* **56**, 615-649 (1987).
- Sullivan, K. A. *et al. Nature* **330**, 758-760 (1987).
- Hirsch, J. P., Dietzel, C. & Kurjan, J. *Genes Dev.* **5**, 467-474 (1991).
- West, R. E., Moss, J., Vaughan, M., Liu, T. & Liu, T. Y. *J. biol. Chem.* **260**, 14428-14430 (1985).
- Simonds, W. F., Goldsmith, P. K., Codina, J., Unson, C. G. & Spiegel, A. M. *Proc. natn. Acad. Sci. U.S.A.* **86**, 7809-7813 (1989).
- Gutowksi, S. *et al. J. biol. Chem.* **266**, 20519-20524 (1991).
- Shenker, A., Goldsmith, P., Unson, C. G. & Spiegel, A. M. *J. biol. Chem.* **266**, 9309-9313 (1991).
- Hamm, H. E. *et al. Science* **241**, 832-835 (1988).
- Wong, Y. H., Conklin, B. R. & Bourne, H. R. *Science* **255**, 339-342 (1992).
- Chabre, O. *et al. Molec. Endocrinol.* **6**, 551-556 (1992).
- Ashkenazi, A. *et al. Science* **238**, 672-675 (1987).
- Laurent, E., Mockel, J., Van Sande, J., Graff, I. & Dumont, J. E. *Molec. cell. Endocrinol.* **52**, 273-278 (1987).
- Cotecchia, S. *et al. J. biol. Chem.* **265**, 63-69 (1990).
- Conklin, B. R., Chabre, O., Wong, Y. H., Federman, A. D. & Bourne, H. R. *J. biol. Chem.* **267**, 31-34 (1992).
- Denker, B. M., Schmidt, C. J. & Neer, E. J. *J. biol. Chem.* **267**, 9998-10002 (1992).
- Woon, C. W., Soparkar, S., Heasley, L. & Johnson, G. L. *J. biol. Chem.* **264**, 5687-5693 (1989).
- Avigan, J. *et al. Biochemistry* **31**, 7736-7740 (1992).
- Dratz, E. A. *et al. Nature* **363**, 276-281 (1993).
- Hamm, H. *Cell molec. Neurobiol.* **11**, 563-578 (1991).
- Okuma, Y. & Reisine, T. *J. biol. Chem.* **267**, 14826-14831 (1992).
- Kleuss, C. *et al. Nature* **353**, 43-48 (1991).
- Cerione, R. A. *et al. J. biol. Chem.* **260**, 1493-1500 (1985).
- Kelleher, D. J. & Johnson, G. L. *Molec. Pharmac.* **34**, 452-460 (1988).
- Phillips, W. J. & Cerione, R. A. *J. biol. Chem.* **267**, 17032-17039 (1992).
- Phillips, W. J., Wong, S. C. & Cerione, R. A. *J. biol. Chem.* **267**, 17040-17046 (1992).
- Kleuss, C., Scherübel, H., Hescheler, J., Schultz, G. & Wittig, B. *Nature* **358**, 424-426 (1992).
- Denker, B. M., Neer, E. J. & Schmidt, C. J. *J. biol. Chem.* **267**, 6272-6277 (1992).
- Strathmann, M. & Simon, M. I. *Proc. natn. Acad. Sci. U.S.A.* **87**, 9113-9117 (1990).
- Sullivan, K. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 6687-6691 (1986).
- Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. *Gene* **77**, 61-68 (1989).
- Guyer, C. A. *et al. J. biol. Chem.* **265**, 17307-17317 (1990).
- Grandy, D. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 9762-9766 (1989).
- Mahan, L. C. *et al. Molec. Pharmac.* **40**, 1-7 (1991).

ACKNOWLEDGEMENTS. We thank S. Coughlin, H. Hamm, F. Cohen, P. Garcia, H. Czerwonka and members of the Bourne laboratory for discussion. This work was supported by grants from the NIH, the March of Dimes, the US-Israel Bi-National Science Foundation, and the NSF Presidential Young Investigator Award. D.J. is a fellow of the McKnight Foundation for the Neurosciences and the PEW Scholar Program.

with the NOE distance constraints. The amino-terminal residues form a turn resembling a single turn of α -helix. Interactions between the amide protons of neighbouring amino acids at the C terminus, such as those seen between Cys 347, Gly 348, Leu 349 and Phe 350 (NOEs between residues 8-9, 9-10, 8-10 and 9-11 in Fig. 2a), provide strong evidence for a β -turn in this region^{18,19}. As β -turns are commonly found in regions of molecular recognition, we tested the importance of this turn for biological activity. Glycine residues are frequently found in the second or third positions of β -turns^{18,19}. Replacement of Gly with any L-amino acid should disrupt the turn formed in the bound state (Fig. 1b legend). But replacement of Gly with the enantiomorphous D-alanine is consistent with the turn structure and is predicted to break other common types of secondary structure. Peptide analogues replacing Gly 348 with D-Ala or L-Leu were synthesized and tested for their ability to stabilize MII. Figure 1b shows that the D-Ala analogue is nearly as potent as the native peptide in stabilizing MII, whereas the L-Leu analogue has greatly reduced activity. The structural and functional data provide evidence for a type II' β -turn^{18,19} between Cys 347 and Phe 350, which is required for peptide stabilization of MII. Alternatively, that interaction with rhodopsin may be sterically blocked by L-amino acid substituents at residue 348.

FIG. 1a, Sequence of the C-terminal 11 amino acids from α_t compared to the deduced amino-acid sequences of other G protein α -subunits^{1,2}. The consensus amino-acid sequence in the C-terminal region of α -subunits is I/LxxNL+xxxLh, where h is a large hydrophobic residue. The conserved leucines at positions 1, 5 and 10 of the sequence and the bulky hydrophobic residue at position 11, suggest that receptor binding sites for the G_q C terminus may be rather hydrophobic. There is evidence for the importance of hydrophobic residues in loops 3 and 4 of the cytoplasmic face of receptors^{29,30}. In the G_i class of G_q subunits (α_t , α_i , α_o and α_2), a glycine is found at position 9 and Met or Asn usually replace this glycine in the other α -subunits. The change of Gly to Asn is proposed to be involved in changing receptor specificity from α_i to α_q ¹⁴. This may be due to a change in the turn geometry of the bound peptide as the Asn side chain can often hydrogen-bond back to the main chain and form a type VIII turn which has a different side-chain presentation from the type II' turn favoured by Gly^{18,19}. b, Analogue peptide stabilization of MII. Increasing concentrations of α_t (340-350) and analogues, or G_i , shift the absorbance spectrum of light-excited rhodopsin from MI (A_{max} at 480 nm) to MII (A_{max} 380 nm)⁹. Rod outer segment (ROS) membranes were washed four times in 10 mM Tris HCl, pH 7.5, 1 mM dithiothreitol and 1 mM EDTA to remove G_i and other peripheral membrane proteins. Absorbance spectra of ROS membranes (5 μ M rhodopsin) in 200 mM NaCl, 10 mM MOPS, pH 8.0, 2 mM $MgCl_2$, 1 mM dithiothreitol and 10 μ M PMSF with increasing concentrations of α_t (340-350) (filled circles), an Arg-substituted analogue, α_t (340-350)K341R (triangles) or analogues at the Gly 348 position (D-Ala open circles; L-Leu, filled squares), or 5 μ M G_i (open square) were measured in the dark at 4 °C with an SLM Aminco DW2000 spectrophotometer. An actinic flash bleached 5% of the rhodopsin, the absorbance spectrum was measured at 1 min, and the light-dark difference spectrum computed. Metarhodopsin II was measured by the difference in absorbance at 380 and 420 nm. An optimum concentration of G_i (5 μ M) produced 4.05×10^{-2} absorbance units of MII. Lack of complete stabilization of MII by peptides may reflect an ability of peptides to bind weakly to the MI precursor as well as to MII. A downturn in initial MII production by the α_t (340-350) peptide at 2 mM compared to 1 mM is reproducible. Light-stimulated rhodopsin is sensitive to nonspecific attack at the Schiff's base linkage of the retinal chromophore by water-soluble amines³¹. The α_t (340-350) sequence has 3 amino groups and is used at millimolar concentrations at which smaller amines attack the retinal Schiff's base. A time- and peptide concentration-dependent decay of MII was seen in the presence of α_t (340-350) ($t_{1/2}$ 93 min at 2 mM). The MII decay was slower in the presence of the α_t (340-350)K341R analogue ($t_{1/2}$ 192 min at 2 mM), and slower still when the N-terminal amino group was acetylated (α_t (Ac-340-350)K341R, $t_{1/2}$ 300 min at 2 mM), suggesting that peptide amino groups can attack the retinal Schiff's base in MII. Conformational energy calculations were carried out, based upon the bound structure of α_t (340-350), to predict the amount of energy needed to form a β -turn around position 348 with other amino acids substituted for Gly. With a relative energy (kcal mol⁻¹) of zero for Gly, the D-Ala analogue was 3.2; L-Ala, 12.4; L-Leu 88.2 without energy minimization. The residual activity

When rhodopsin-peptide mixtures are exposed to light to form photoexcited MII, the Tr-NOESY crosspeak intensities of the α_t (Ac-340-350)K341R peptide analogue show substantial changes (Fig. 2c-f). NMR measurements on the light-excited samples were run for 2-4 h, consistent with the lifetimes of the MII-peptide complexes (Fig. 1b legend). The shorter experiment times lead to lower signal-to-noise Tr-NOESY spectra than were obtained for the rhodopsin-bound peptide. About 30 new Tr-NOESY peaks and about 30 substantial changes in NOESY intensity are observed if the bleached samples are compared to the pre-bleached samples as shown in Fig. 2c-f under the same low signal-to-noise conditions. Many of the side chains have more defined conformation in the MII-bound peptide, as shown by new intraresidue NOEs (Fig. 2c-h). The approximate MII bound structure that resulted from constrained molecular dynamics is shown in Fig. 3b and d. The helix-like tight turn at the N terminus becomes more relaxed. Lys 345, which is close to Glu 342 in the rhodopsin-bound peptide, now appears to be closer to the free carboxyl at the C terminus (compare Fig. 3c and d). A salt bridge between Glu 342 and Lys 345 may have been broken by the activated receptor and replaced by another between Lys 345 and the C-terminal carboxyl, favouring relaxation of the N-terminal turn. A number of new short- and medium-



observed with the L-Leu analogue may indicate that the N-terminal turn in the peptide helps stabilize the receptor-bound conformation, and the reduced activity of N-terminal-truncated peptide analogues supports this (data not shown). Data represent mean $\pm$ s.e.m. ($n=4$) for experiments which were repeated at least 3 times. Error bars are within the area of the symbols.

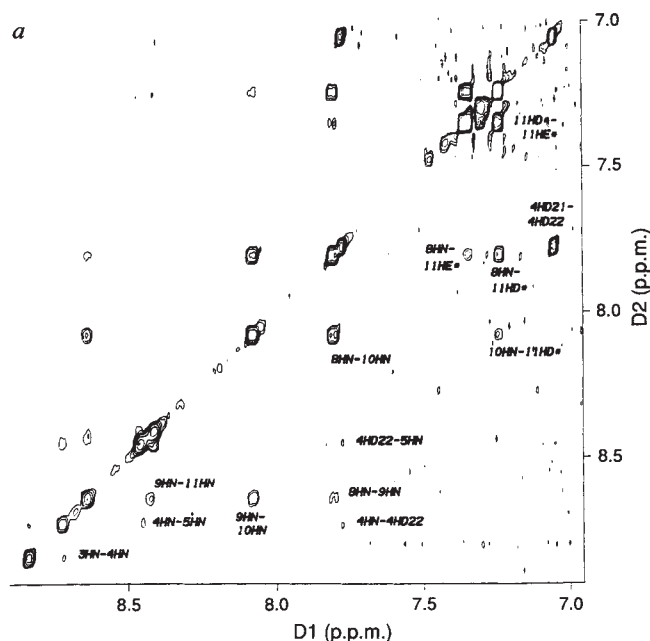
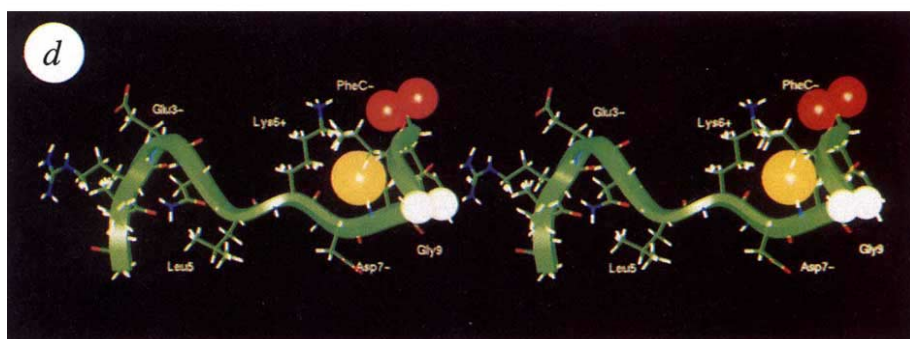
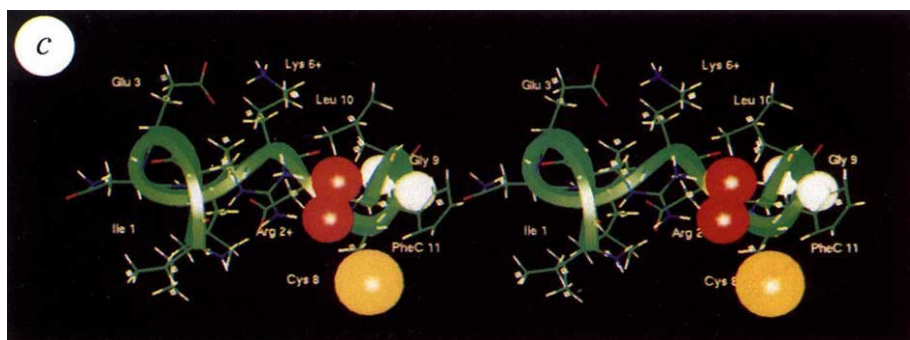
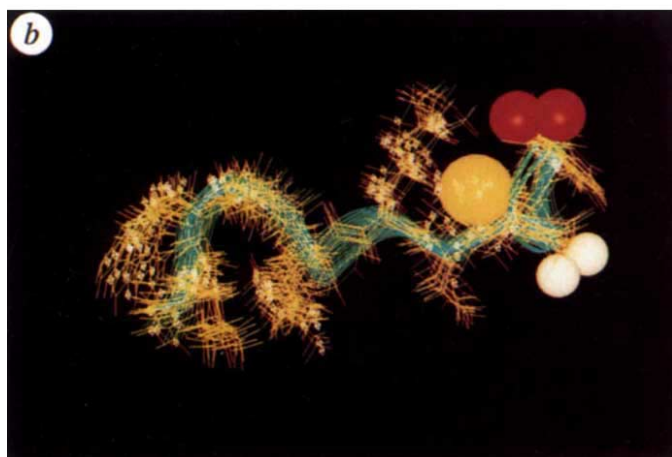
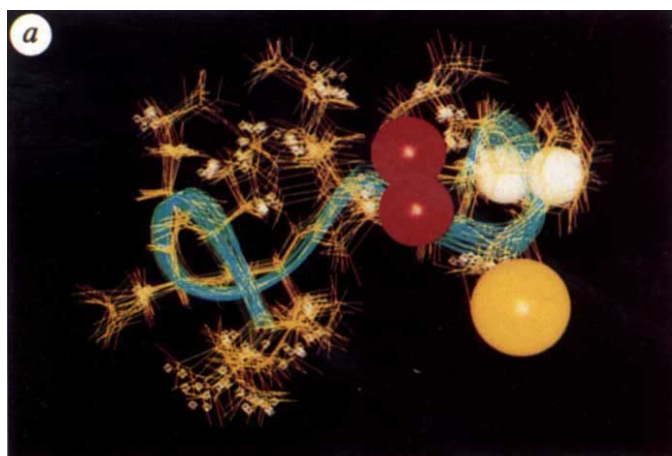


FIG. 3a, Results of constrained molecular dynamics of rhodopsin-bound $\alpha_1(340-350)$ K341R. Several different starting conformations were used for the constrained molecular dynamics and simulated annealing, including extended, β -sheet, right-handed α -helix, left-handed α -helix and distance geometry structures. Charges were omitted as the dielectric environment of the receptor binding site is unknown (but probably rather hydrophobic; see Fig. 1a legend). Structures were archived every 0.1–1 ps and energy minimized using steepest descents followed by conjugate gradients. An assessment of the degree to which the structures have been constrained can be obtained from differences between overlays of the dynamics results for the main chain and the side chains. The r.m.s. deviation for the main chain atoms is about 1.0 Å. Many of the side chains and the N-terminal region are less well defined and the uncertainties may be due to thermal motion and/or undetermined distance constraints. Some of the most important atoms in the structure have been highlighted with full diameter space-filling atoms. The Cys sulphur, shown as the large yellow ball, is the pertussis toxin ADP-ribosylation site. The negative oxygens of the C-terminal carboxyl which are

essential for stabilization of the active form of the receptor are shown as a pair of red balls. The Gly 348 hydrogens that are located on a β -turn are shown as two white balls. NOEs were first divided into four classes, short (<2.3 Å), medium (<2.7 Å), long (<3.5 Å) and very long (<4.5 Å). Pseudotomographs were used for pairs of protons that were not stereospecifically assigned. Constrained molecular dynamics was run at high temperatures (500 or 1,000K) for times up to 1,000 ps using the program DISCOVER (Biosym) and the CVFF forcefield. Peptide bonds were maintained in the *trans* conformation by increasing the force constant for rotation about the peptide bond. A grid search of the ϕ/ψ angles in the C-terminal half of the dark rhodopsin-bound structure was carried out with calculations of NOE and energy agreement to ensure that the dynamics did not get trapped in false minima. *b*, Results of constrained molecular dynamics of Mil-bound α_1 (Ac-340–350)K341R. The intensity changes and new peaks in the Mil-bound peptide were used to modify the distance constraints in the rhodopsin-bound structure and molecular dynamics and simulated annealing were run as described. A number of new short- and medium-range interactions are observed between the amino-terminal 5 amino acids. *c*, Stereo pair of a selected structure from *a* of the rhodopsin-bound α_1 (Ac-340–350)K341R. A representative structure that showed good agreement with the NOE constraints. *d*, Stereo pair of a selected structure from *b* of the Mil-bound α_1 (Ac-340–350)K341R.

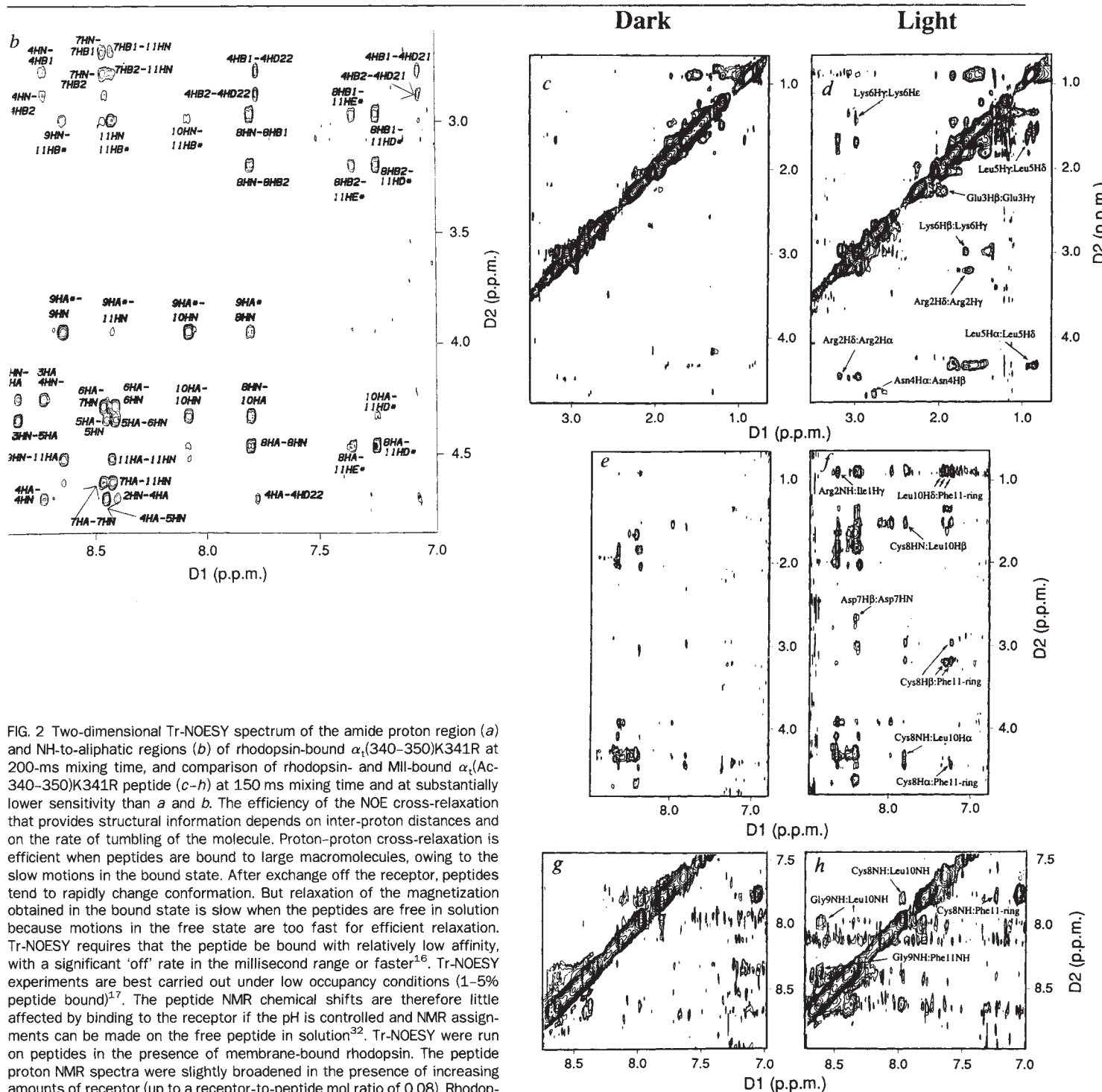
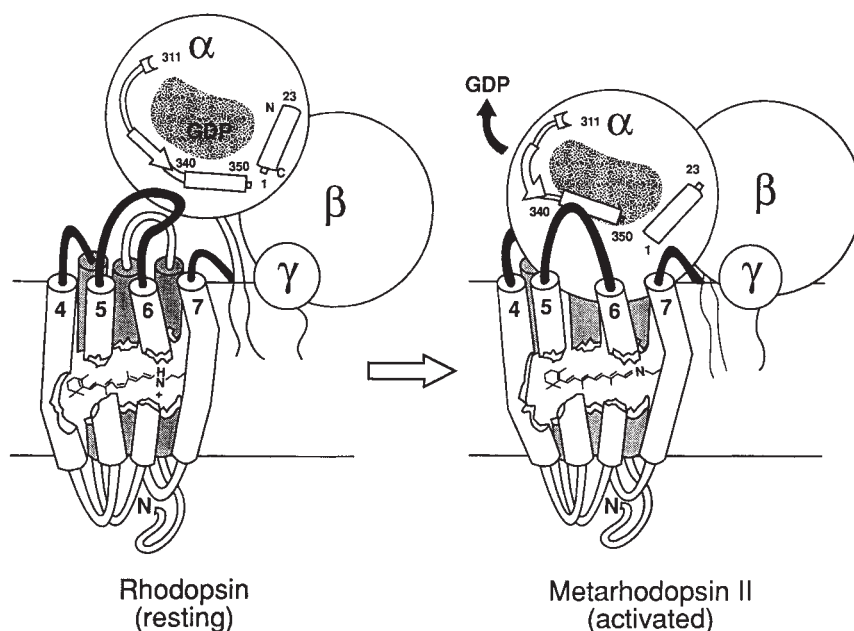


FIG. 2 Two-dimensional Tr-NOESY spectrum of the amide proton region (a) and NH-to-aliphatic regions (b) of rhodopsin-bound $\alpha_1(340-350)$ K341R at 200-ms mixing time, and comparison of rhodopsin- and MII-bound $\alpha_1(\text{Ac-340-350})$ K341R peptide (c-h) at 150 ms mixing time and at substantially lower sensitivity than a and b. The efficiency of the NOE cross-relaxation that provides structural information depends on inter-proton distances and on the rate of tumbling of the molecule. Proton-proton cross-relaxation is efficient when peptides are bound to large macromolecules, owing to the slow motions in the bound state. After exchange off the receptor, peptides tend to rapidly change conformation. But relaxation of the magnetization obtained in the bound state is slow when the peptides are free in solution because motions in the free state are too fast for efficient relaxation. Tr-NOESY requires that the peptide be bound with relatively low affinity, with a significant 'off' rate in the millisecond range or faster¹⁶. Tr-NOESY experiments are best carried out under low occupancy conditions (1-5% peptide bound)¹⁷. The peptide NMR chemical shifts are therefore little affected by binding to the receptor if the pH is controlled and NMR assignments can be made on the free peptide in solution³². Tr-NOESY were run on peptides in the presence of membrane-bound rhodopsin. The peptide proton NMR spectra were slightly broadened in the presence of increasing amounts of receptor (up to a receptor-to-peptide mol ratio of 0.08). Rhodopsin concentration was up to 9 mg ml⁻¹ (450 μ M), the peptide concentration was 3-20 mM in 20 mM phosphate, pH 7.5, 100 mM KCl, 0.1 mM EDTA, 1 mM DTT, 5% D₂O/95% H₂O, with DSS as internal standard; temperature was 1 °C. The peptide did not bind detectably to lipid bilayers, no Tr-NOESY signals being obtained from the peptide in the presence of liposomes made of retinal rod disk lipids³³ (data not shown). NMR measurements were made as described³². Stereospecific assignment awaits structure refinement. Tr-NOESY crosspeak intensities of the $\alpha_1(340-350)$ peptide with retinal rod membranes were studied as a function of the mixing time. At sufficiently low mixing times, Tr-NOESY intensities are approximately related to the inverse sixth power of the distance between protons¹⁷. The intensities of NOESY peaks were integrated using FELIX (Biosym). Interproton distances were calibrated using intensities of crosspeaks between protons in the peptide which do not depend on conformation, including Cys β - β (1.75 Å) and Asn δ - δ (1.80 Å). In some data sets, the Phe ring 2-3/5-6 distance (2.46 Å) could also be used. If mixing times were sufficiently short (<150 ms), there was agreement between the different distance calibrations. 168 NOEs were measured in the rhodopsin-bound peptide, 118 of

which provided structurally significant proton-proton distance constraints in the rhodopsin-bound peptide. c-h, Comparison of dark and light Tr-NOESY on $\alpha_1(\text{Ac-340-350})$ K341R at the same scale in different regions of the data set. ROS membranes were washed in the buffer used for the dark experiments, adjusted to pH 6.5. Peptides were at 6.5 mg ml⁻¹ and adjusted to pH 6.5; rhodopsin concentration was 6 mg ml⁻¹. After NMR measurements in the dark, samples were bleached at 0 °C with a 100-watt quartz iodine lamp for 2 min and a 520-nm-long wavelength pass filter. In the light, the side chains are more immobilized (c versus d), the NH protons have much better defined interactions with the α , β and methyl protons of other side chains (e versus f), and many crosspeaks in the NH-NH region are substantially increased (g versus h). Because of the shorter measuring times (3-4 h), the Tr-NOESY data had a much lower signal-to-noise ratio on the light-exposed samples. There were fewer MII-bound peptide distance constraints and the constraints were not as narrowly defined as for the dark rhodopsin bound structure. There were 73 interproton distance constraints in the MII-bound peptide, 57 intraresidue, 10 sequential, 3 with $\Delta i=2$ and 3 with $\Delta i=3$.

FIG. 4 Proposed model of receptor-mediated G-protein activation. A relatively low-affinity interaction between inactive rhodopsin and G_i is indicated by ability of the carboxyl terminal of α_i to bind to dark rhodopsin detected by Tr-NOESY. Light activation of rhodopsin opens up additional sites of interaction and there is a large increase in affinity between the two proteins (activated state). The N- and two C-terminal regions of α_i (ref. 9) are potential rhodopsin interaction sites. The three filled regions on rhodopsin's cytoplasmic surface that block interaction with G_i (ref. 7) are shown as complementary interacting regions. The MII-mediated conformational change around Lys 345 and opening of the β -turn structure at the α_i carboxy terminus is shown schematically in the diagram. This conformational change is proposed to play a role in decreasing GDP affinity for the nucleotide binding site on α_i , leading to GDP release and G-protein activation. GTP binding to the empty nucleotide site causes a conformational change on α_i , resulting in a decreased affinity for receptor and $\beta\gamma$ subunits, and effector activation.



range NOE interactions are observed between the amino-terminal five amino acids. The Cys side chain appears to pivot from pointing 'outside' the rhodopsin-bound peptide to being tucked inwards in the MII bound structure (Fig. 3c, d), and there appears to be an opening of the turn conformation at the C terminus.

An interesting paradox is raised by these data. Analogue peptide studies are consistent with a β -turn around Gly 348 being required for MII stabilization. Mutagenesis showing that three of the last four terminal amino acids are important in specifying receptor selectivity is consistent with this result¹⁴. However, the β -turn no longer seems to be present in the MII-bound peptide (Fig. 3d). We suggest that this may be because the activated receptor first binds to and then breaks the β -turn as part of the G-protein-activation process. This speculation is testable using mutant rhodopsins that bind G_i with high affinity but no longer activate it²⁰. The Tr-NOESY of peptides bound to the mutant MII should not have the light-activated conformation.

The data suggest a structural mechanism by which pertussis toxin could uncouple substrate G proteins from their cognate receptors^{21,22}, because insertion of an ADP-ribose group would be predicted to change the conformation of the C terminus. Our data predict that any L-amino acid at position -3 would disrupt the C-terminal β -turn, which might explain results from mutagenesis in which substitution of Gly 353 in α_o corresponding to Gly 348 in α_i , yields a G protein that is no longer recognized by pertussis toxin²³.

Switching of receptor specificity by changing three of the terminal four amino acids of α_q for α_i/α_o sequences¹⁴ suggests that subfamilies of β -turns may be involved in specifying receptor selectivity, because the α_q sequence is predicted to form a different type of β -turn^{18,19} from the α_i/α_o type of sequences (Fig. 1a legend).

The conformational changes inferred in the approximate MII-bound structure include reorganization of both the amino- and carboxyl-terminal turns of the peptide. Flexibility of the peptide bond around Lys 345 of α_i may be important for adopting the productive, MII-bound structure. The *unc* mutation, from the corresponding Arg 389 to the more rigid Pro in α_{11} disallows receptor- G_s coupling, providing *in vivo* evidence that a conformational change in this region may be functionally significant.

Stimulated receptors activate G proteins by catalysing GDP release, the rate-limiting step in GTP/GDP exchange. Interestingly, a model of the GDP-binding pocket of α_i places Lys 345

near the GDP-binding site, based upon homology with the GDP-binding pocket of EF-Tu^{24,25}. Mutagenesis²⁶ has shown that the C-terminal ten amino acids of the α -subunit are important in determining GDP affinity in G proteins. Therefore a receptor-mediated change in the position of Lys 345 and surrounding amino acids may be a determinant of decreased GDP affinity and lead to G-protein activation, as in the model of the receptor-G-protein interaction shown in Fig. 4. Interestingly, mutants in topologically similar regions of yeast Ras and bacterial EF-Tu are able to overcome the need for exchange factor-mediated catalysis of GTP/GDP exchange^{27,28}.

Our approach could be extended to the structures of the active interfaces mediating other regulatory or adhesive protein-protein interactions. This information should guide the synthesis of locked conformation or non-peptide mimetic compounds with higher affinity. Such compounds have the potential to form agonists or antagonists of specific receptor-protein interactions. □

Received 17 December 1992; accepted 25 March 1993.

- Simon, M. I., Strathmann, M. P. & Gautam, N. *Science* **252**, 802-808 (1991).
- Kaziro, Y., Itoh, H., Kozasa, T., Nakafuku, M. & Satoh, T. A. *Rev. Biochem.* **60**, 349-400 (1991).
- Dohlman, H. G., Thorner, J., Caron, M. G. & Lefkowitz, R. J. A. *Rev. Biochem.* **60**, 653-688 (1991).
- Birnbaumer, L. A. *Rev. Pharmac. Toxicol.* **30**, 675-705 (1990).
- Hamm, H. E. *Cell molec. Neurobiol.* **11**, 563-578 (1991).
- Hargrave, P. A., Hamm, H. E. & Hofmann, K. P. *Bioassays* **15**, 1-8 (1993).
- Koenig, B. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 6878-6882 (1989).
- Hausdorf, W. P., Hnatowich, M., O'Dowd, B. F., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **265**, 1388-1393 (1990).
- Hamm, H. E. *et al. Science* **241**, 832-835 (1988).
- Spiegel, A. M., Simonds, W. F., Jones, L. F., Goldsmith, P. K. & Unson, C. G. *Soc. gen. Physiol. ser.* **45**, 185-195 (1990).
- Sullivan, K. A. *et al. Nature* **338**, 758-762 (1987).
- Sukumar, M. & Higashijima, T. *J. biol. Chem.* **267**, 21421-21424 (1992).
- Weingarten, R., Ransnas, L., Mueller, H., Sklar, L. A. & Bokoch, G. M. *J. biol. Chem.* **265**, 11044-11049 (1990).
- Conklin, B. R., Farfel, Z., Lustig, K. D., Julius, D. & Bourne, H. R. *Nature* **363**, 274-276 (1993).
- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **348**, 125-132 (1990).
- Clore, G. M. & Gronenborn, A. M. *J. magn. Reson.* **53**, 423-442 (1983).
- Campbell, A. P. & Sykes, B. D. *J. magn. Reson.* **93**, 77-92 (1991).
- Wilmot, C. M. & Thornton, J. M. *J. molec. Biol.* **203**, 221-232 (1988).
- Wilmot, C. M. & Thornton, J. M. *Protein Engng* **3**, 479-493 (1990).
- Frankie, R. R., Koenig, B., Sakmar, T. P., Khorana, H. G. & Hofmann, K. P. *Science* **250**, 123-125 (1990).
- Van Dop, C. *et al. J. biol. Chem.* **259**, 23-26 (1984).
- Kurose, H., Katada, T., Amano, T. & Ui, M. *J. biol. Chem.* **258**, 4870-4875 (1983).
- Avigan, J. *et al. Biochemistry* **31**, 7736-7740 (1992).
- Deretic, D. & Hamm, H. E. *J. biol. Chem.* **262**, 10839-10847 (1987).
- Jurnak, F. *Science* **230**, 32-36 (1985).
- Hwang, Y.-W., Carter, M. & Miller, D. L. *J. biol. Chem.* **267**, 22198-22205 (1992).
- Denker, B. M., Schmidt, C. J. & Neer, E. J. *J. biol. Chem.* **267**, 9998-10002 (1992).
- Camonis, J. H. & Jacquet, M. *Molec. cell. Biol.* **8**, 2980-2983 (1988).
- Cheung, A. H., Huang, R. C. & Strader, C. D. *Molec. Pharmacol.* **41**, 1061-1065 (1992).
- Dixon, R. A. F., Sigal, I. S. & Strader, C. D. *Cold Spring Harbor Symp. Quant. Biol.* **LIII**, 487-497 (1988).

31. Hubbard, R., Brown, P. K. & Bownds, D. *Meth. Enzym.* 615–653 (1971).
 32. Ostheimer, G. J., Starkey, J. R., Lambert, C. G., Helgeson, S. L. & Dratz, E. A. *J. Biol. Chem.* 267, 120–128 (1992).
 33. Brown, M. F., Deese, A. J. & Dratz, E. A. *Meth. Enzym.* 81, 709–728 (1982).

ACKNOWLEDGEMENTS. We thank N. Artemyev, J. Mills, N. Skiba, M. Tientze and A. Jesaitis for discussion; P. Sigler, H. Bourne and M. Rasenick for critically reading the manuscript; and C. Johnson and T. Garcia for technical assistance. Research supported by grants from the NIH, AHA, Glaxo, NSF, and Montana Center for Excellence in Biotechnology. H.M.R. was supported by an NIH training grant.

Specific repression of TATA-mediated but not initiator-mediated transcription by wild-type p53

David H. Mack^{*†}, Jai Vartikar[‡], James M. Pipas[‡] & Laimonis A. Laimins^{*§}

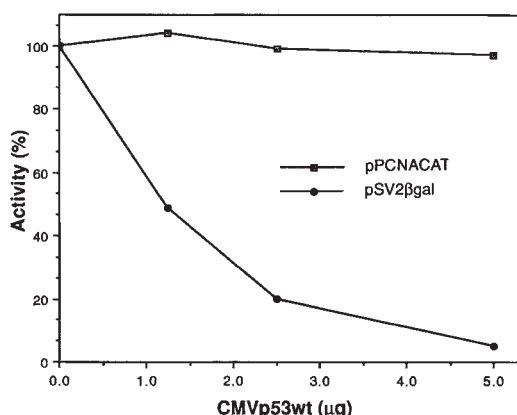
^{*} Department of Molecular Genetics and Cell Biology, [§] Howard Hughes Medical Institute, The University of Chicago, 5841 S. Maryland Avenue, MC1028, Chicago, Illinois 60637, USA

[‡] Department of Biological Sciences, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA

[†] Present address: Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA

^{||} To whom correspondence should be addressed

THE p53 protein is apparently central to the development of human cancers because both alleles are often found to be mutated in different tumour types¹. In addition, wild-type p53 can inhibit transformation by viral and cellular oncogenes *in vitro*, so p53 has been classified as a tumour suppressor². Investigations of the normal function of p53 have indicated that at least one of its functions could involve the activation of gene expression through the binding of specific DNA-regulatory sequences^{3,4}. Also, overexpression of p53 can mediate growth arrest⁵ and repress transcription from a variety of promoters^{6,7}. We demonstrate here both *in vivo* and *in vitro* that expression of wild-type p53 specifically represses the activity of promoters whose initiation is dependent on the presence of a TATA box. Promoters whose accurate transcription is directed by a pyrimidine-rich initiator element, however, are immune to the effects of p53. Furthermore, we observe that repression is mediated by an interaction of p53 with basal transcription factor(s). Thus, p53 appears to repress the activity of certain promoters through direct communication with TATA box-dependent basal transcription machinery.



We have previously reported that co-transfection of an expression vector encoding human wild-type p53 with various reporter constructs, including pSV2CAT, pBLCAT2 and p56fos-CAT (using only the *fos* TATA box to drive expression) results in a dose-dependent repression of transcription⁷. In contrast, co-transfection with plasmids encoding mutant p53 proteins that fail to suppress transformation in tissue culture assays has no effect on reporter expression. Of the constructs tested, we found that the proliferating cell nuclear antigen (PCNA) promoter was resistant to the repressive effects of p53, whereas in the same transfection reaction transcription directed by the SV40 early region was efficiently inhibited (Fig. 1). Inclusion in our analysis of two detectable enzymatic activities in a single transfection demonstrated that p53 effects on expression in these experiments were not a secondary consequence of growth arrest. Similar results were observed in human cervical cells (HeLa), mouse NIH 3T3 fibroblasts and the human lymphocyte cell line Jurkat (not shown). In addition to PCNA, other promoters have been reported to be resistant to repression by p53 in transient assays⁸. These include the *c-Ha-ras* and the epidermal growth factor receptor promoters. Interestingly, all three of the promoter sequences unaffected by wild-type p53 do not use TATA elements to direct transcription, but use initiator elements instead. In contrast, all TATA-containing promoters we have tested so far have been sensitive to repression by p53.

We compared the sensitivity to the repressive effects of p53 of promoters using TATA-containing elements with those that use initiator elements with the aid of two synthetic promoter constructs. Both contained identical 21-base-pair (bp) upstream activator elements, but they differed downstream in having either an adenovirus major late promoter TATA box or the terminal transferase gene initiator element to direct the initiation of transcription. Co-transfection with a p53 expression vector resulted in efficient dose-responsive repression of TATA-mediated transcription, whereas initiator-mediated expression was relatively unaffected (Fig. 2). It thus appeared that inhibition of expression by p53 was directed through the particular sequence motif responsible for the accurate initiation of transcription.

To confirm that the specific repressive effects of p53 were indeed mediated at the level of transcription initiation, transient expression was assayed with a plasmid that contained a tandem TATA and initiator element downstream from the SV40 21-bp repeats (Fig. 3a). In the absence of co-expressed p53, transcripts were found by primer extension analysis to initiate predominantly at sites directed by TATA sequences, with a lower but detectable amount initiating at initiator sites. A dominance of TATA-mediated over initiator-mediated transcription has been noted previously from promoters that contain both ele-

FIG. 1 PCNA promoter activity is unaffected by wild-type p53. The human cervical epithelial cell line C33A (ref. 22) was transiently transfected with both PCNACAT and pSV2βgal reporter constructs, together with increasing amounts of an expression vector encoding the wild-type human p53 protein. The ordinate is the per cent change in chloramphenicol acetyltransferase (CAT) or β-galactosidase (βgal) activity due to co-transfection of the p53 expression construct at the amount indicated on the abscissa. The results shown are from a single experiment. Although the degree of change varied slightly between experiments, the relative effects of p53 on the reporter constructs are consistent in 4 different experiments. Similar results were observed in the mouse fibroblast cell line NIH 3T3, and the human lymphocyte cell line Jurkat (data not shown). Repression of the pSV2βgal construct was not specific to the reporter sequences as similar results were obtained with pSV2CAT plasmids⁷. METHODS. PCNACAT²³, pSV2βgal (Pharmacia), and pCMVp53wt (pC53-SN3)⁵ have been described. Transfections in the C33A line were done using the calcium phosphate method²⁴ and contained 5 μg of each reporter plasmid with the indicated amount of p53 effector construct. The total quantity of DNA was brought to 30 μg per 100-mm plate with expression vector lacking p53 sequences. 40 hours post-transfection, cells were extracted and assayed for CAT and β-gal activity²⁴. All assays were normalized for protein concentration.

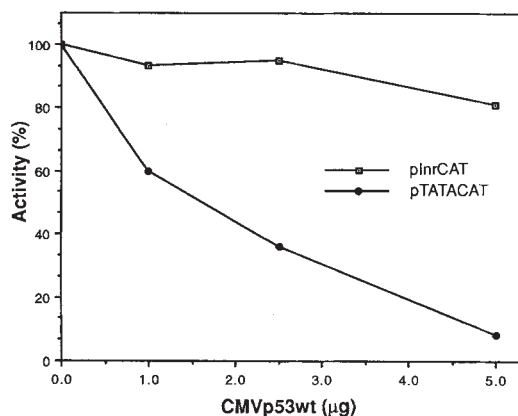


FIG. 2 Initiator (Inr)-mediated transcription is refractory to repressive affects of p53. The effect of increasing amounts of p53 expression plasmid on the activities of two synthetic Sp1-dependent promoters was compared. plnrCAT and pTATACAT reporters differ only in that the terminal transferase initiator element is responsible for directing transcription from the former, whereas the AdML TATA box is used by the latter. The ordinate is the per cent change in CAT activity of reporter plasmid due to cotransfection of p53 expression plasmid at the amounts indicated from a single experiment. Although there was some variation in the degree of repression between individual experiments, relative activities were consistent in four separate experiments. METHODS. Construction of pTATACAT and plnrCAT was achieved by subcloning a *Bam*HI-*Hinc*II fragment from plasmid V (ref. 9) containing the SV40 21-bp repeats and the AdML TATA, or from plasmid VI (ref. 9) containing the SV40 21-bp repeats and the TdT Inr, respectively, into the *Bgl*III site of the pCAT3M²⁵. Incompatible ends were blunt-ended with T4 polymerase and ligated. Reporter plasmid (10 μg) was transfected with increasing amounts of p53 expression construct into C33A cells and CAT was assayed as described in Fig. 1 legend.

ments⁹. In the presence of co-transfected p53 expression vector, transcripts initiated by the TATA box decreased but those directed by the initiator element were relatively unchanged (Fig. 3b). At high concentrations of co-transfected p53 plasmid, there was a slight increase in initiator-mediated transcription, which was probably the result of reduced promoter competition with TATA-initiated transcription. Taken together, these experiments demonstrate that p53's effect on transcription is not due to transcriptional 'squenching'¹⁰ or to a general nonspecific shut-down of RNA polymerase II machinery.

The observed repressive effects of high levels of p53 protein on expression of TATA-containing promoters could be due either to a direct effect of p53 on the basal transcription machinery or to an indirect effect mediated through a master repressor(s) which might be induced as a result of p53's sequence-specific transactivating properties^{3,4}. To discriminate between these two possibilities, we tested whether p53 could mediate repression of transcription *in vitro*. Using a template with both TATA and initiator element sequences downstream from the Sp1 activator element in transcription reactions *in vitro* with HeLa cell nuclear extracts, we found the TATA to be more efficient than the initiator element at initiating transcription (Fig. 4). As we found *in vivo*, addition of purified wild-type mouse p53 protein resulted in specific repression of TATA-initiated

transcripts but initiator-directed transcription was unaffected or even slightly transactivated. At very high concentrations of purified p53, all transcriptional activity was inhibited, probably as a result of nonspecific 'squenching' effects¹⁰ (data not shown).

The specific repression we observe *in vitro* is consistent with a direct interaction of p53 with TATA-dependent basal transcription machinery. Interestingly, both TATA and initiator elements use TATA-binding protein (TBP) for their activities¹¹, but each relies on a different complement of TBP-associated factors (TAFs) for function¹². Hence it is likely that repression by p53 is mediated either through an interaction with a basal transcription component or TAF, which results in an interference with the formation of an essential TATA transcription complex. A number of nuclear factors that associate with TBP and result in a functional repression of transcription have been described¹³⁻¹⁵ and, consistent with our observations, p53 has been shown to bind TBP directly¹⁶.

Although our experiments have demonstrated that transcription from promoters differing only in the type of transcriptional initiating motif present are differentially affected by p53, it is unlikely that all TATA box promoters will be affected to the same degree. Given our results, the particular composition of the transcription initiation complex on the promoter is likely to influence its sensitivity to p53. For example, TFIIE is required

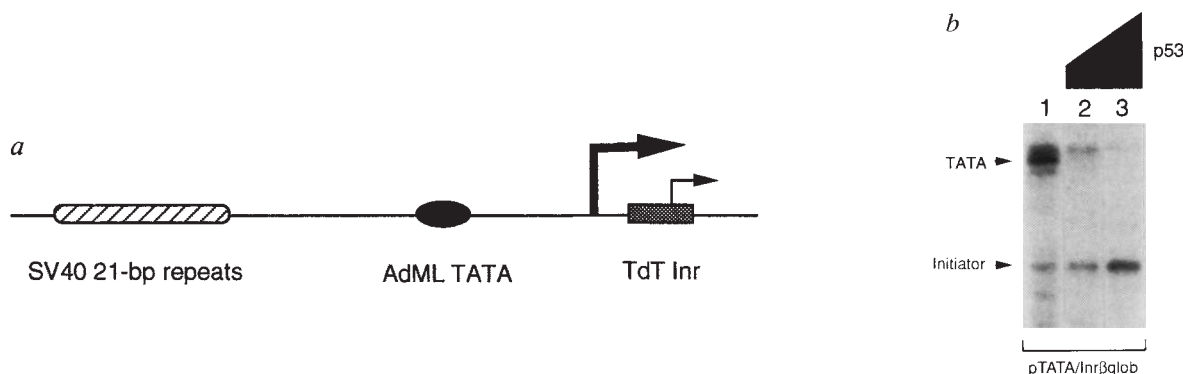


FIG. 3 p53 specifically represses TATA-motif-directed transcription *in vivo*. a, Promoter containing both TATA and initiator (Inr) elements separated by 35 bp, downstream of the SV40 21-bp repeats. Large and small arrows indicate the approximate locations of the transcription start sites for AdML TATA box and TdT Inr respectively. This promoter was used to drive expression of the rabbit β -globin gene in plasmid pTATA/Inr β glob. b, Primer extension analysis of transfected pTATA/Inr β glob activity with increasing amounts of p53 expression vector in C33A cells. Primer extension products of transcripts that initiate at TATA and Inr are indicated. Similar results were obtained from HeLa cells and mouse NIH 3T3 cells (data not shown).

METHODS. pTATA/Inr β glob was constructed by inserting the *Sal*I fragment of plasmid IX (ref. 9) containing the SV40 21-bp repeats upstream of the AdML TATA box and TdT Inr in SP6, into *Sal*I-*Pst*I restricted Ovec-1 (ref. 26). Incompatible ends were blunt-ended with T4 polymerase and ligated. 10 μg of reporter construct was co-transfected with: 0 μg (lane 1), 2.5 μg (lane 2), or 5.0 μg (lane 3) of pCMVp53wt plasmid. Transfections were done as described in Fig. 1 legend. Cytoplasmic RNA was isolated 40 h post-transfection for primer extension analysis using a primer derived from the rabbit β -globin gene, followed by electrophoresis of labelled cDNA products on an 8% denaturing polyacrylamide gel.

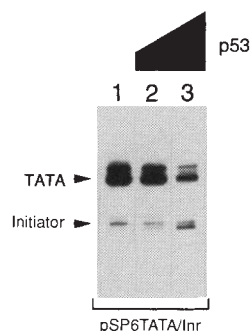


FIG. 4 Effects of p53 on TATA- and 'initiator' (Inr)-mediated transcription *in vitro*. Purified p53 protein specifically inhibits TATA-directed transcription in HeLa cell nuclear extracts. Quantitation by densitometry of primer extension products is as follows: lane 1, TATA (100%)/Inr (100%); lane 2, TATA (73%)/Inr (116%); lane 3, TATA (23%)/Inr (183%).

METHODS. Construction of pSP6TATA/Inr has been described (plasmid IX in ref. 9). Purified wild-type murine p53 protein: 0 ng (lane 1), 50 ng (lane 2), or 200 ng (lane 3) was added to *in vitro* transcription reaction mixtures and preincubated for 10 min at 30 °C. 100 ng pSP6TATA/Inr template and nucleoside triphosphates were added to the reactions and incubations continued for 30 min at 30 °C. RNA transcripts purified from the reaction mixtures were analysed by primer extension with a SP6 promoter primer (Promega) and electrophoresed as described in Fig. 3 legend.

for AdML core promoter activity but is dispensable for accurate initiation of transcription from the immunoglobulin heavy-chain core promoter¹⁷. Although both these elements use TATA boxes for initiation, p53 may have varied effects on each, given the difference in their requirements for particular basal transcription factors.

We observe repressive effects by p53 on transcription only when p53 is overexpressed. The physiological relevance of this may be revealed *in vivo* when p53 is induced to accumulate over normal levels, for example in cells treated with γ -irradiation, ultraviolet light or radiomimetic drugs, in which high levels of wild-type p53 accumulate through a post-translational stabilization mechanism^{18,19}. This rapid accumulation of p53 mediates the temporary arrest of cells at G1, possibly allowing for repair of damaged DNA before entry into S phase¹⁸. This phenomenon is similar to the growth-inhibitory effects of p53 overexpression in culture which result in arrest of the cell cycle at the G1/S boundary²⁰. We would suggest that a shut-down of a particular class of gene expression by increased levels of p53 could be the mechanism by which p53-mediated G1 arrest occurs. Interestingly, many housekeeping genes, as well as those involved in DNA repair, such as PCNA²¹, use initiator elements for transcription initiation. These promoters would be immune to repression by p53 and so expression would continue during periods of environmental stress. Our results therefore indicate that p53 may be important as a negative regulator of transcription during the dramatic effects brought about by DNA damage. □

11. Pugh, B. F. & Tjian, R. *Genes Dev.* **5**, 1935-1945 (1991).
12. Pugh, B. F. & Tjian, R. *Cell* **61**, 1187-1197 (1990).
13. Inostroza, J. A., Mermelstein, F. H., Ha, I., Lane, W. S. & Reinberg, D. *Cell* **70**, 477-489 (1992).
14. Meisterernst, M. & Roeder, R. G. *Cell* **67**, 557-567 (1991).
15. Meisterernst, M., Roy, A. L., Lieu, H. M. & Roeder, R. G. *Cell* **66**, 981-993 (1991).
16. Seto, E. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 12028-12032 (1992).
17. Parvin, J. D., Timmers, H. T. M. & Sharp, P. A. *Cell* **68**, 1135-1144 (1992).
18. Maltzman, W. & Czyzyk, L. *Molec. cell. Biol.* **4**, 1689-1694 (1984).
19. Kastan, M. B., Onyekwere, O., Sidransky, D., Vogelstein, B. & Craig, R. W. *Cancer Res.* **51**, 6304-6311 (1991).
20. Martinez, J., Georgoff, I., Martinez, J. & Levine, A. J. *Genes Dev.* **5**, 151-159 (1991).
21. Nichols, A. F. & Sancar, A. *Nucleic Acids Res.* **20**, 2441-2446 (1992).
22. Scheffner, M., Münger, K., Byrne, J. C. & Howley, P. M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5523-5527 (1991).
23. Morris, G. F. & Mathews, M. B. *J. Virol.* **65**, 6397-6406 (1991).
24. Gorman, C. in *DNA Cloning* (ed., Glover, D. M.) Vol. II, 143-159 (IRL, Oxford, 1985).
25. Gorman, C., Moffat, L. F. & Howard, B. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
26. Westin, G., Gerster, T., Müller, M. M., Schaffner, G. & Schaffner, W. *Nucleic Acids Res.* **15**, 6787-6798 (1987).

ACKNOWLEDGEMENTS. We thank S. Smale and M. Mathews for plasmids. This work was supported by a grant to L.A.L. from the National Cancer Institute and the Howard Hughes Medical Institute.

Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing

Jamal Tazi*, Ute Kornstädt†, Ferdinand Rossi*, Philippe Jeanteur*, Guy Cathala*, Claude Brunel* & Reinhard Lührmann†

* UA CNRS 1191, Génétique Moléculaire Université Montpellier II, Sciences et Techniques du Languedoc, 34095 Montpellier cedex 5, France

† Institut für Molekularbiologie und Tumorforschung, Philipps-Universität Marburg, Emil-Mannkopff-Strasse 2, D-3550 Marburg, Germany

THE U1 small nuclear ribonucleoprotein (snRNP) particle is one of the Sm class of snRNPs essential for splicing of precursor messenger RNA¹⁻⁵. Mammalian U1 snRNP contains a 165-nucleotide long RNA molecule and at least 11 proteins: the U1-specific 70K proteins A and C, and the common U snRNP proteins (B', B, D1, D2, D3, E, F and G). One of the functions of U1 snRNP is recognition of the 5' splice site, an event that requires both U1 RNA and U1 proteins⁶⁻¹⁰. The 70K protein is the only heavily phosphorylated U1 protein in the cell^{11,12}. Isolated U1 snRNPs are associated with a kinase activity that selectively phosphorylates the 70K protein *in vitro* in a reaction requiring ATP. Here we investigate the role of phosphorylation of the 70K protein in the splicing of pre-mRNA. The 70K protein on U1 snRNPs was phosphorylated *in vitro* with either ATP, or with ATP- γ S, which gave a thiophosphorylated product that was resistant to dephosphorylation by phosphatases. When HeLa nuclear splicing extracts that had been depleted of endogenous U1 snRNPs were complemented with U1 snRNPs possessing normal phosphorylated 70K protein, mature spliceosomes were generated and the splicing activity of the extracts was fully restored. By contrast, if thiophosphorylated U1 snRNPs were used instead, splicing was completely inhibited, although formation of the mature spliceosome was unaffected. Our data show that the state of phosphorylation of the U1-specific 70K protein is critical for its participation in a pre-catalytic step of the splicing reaction.

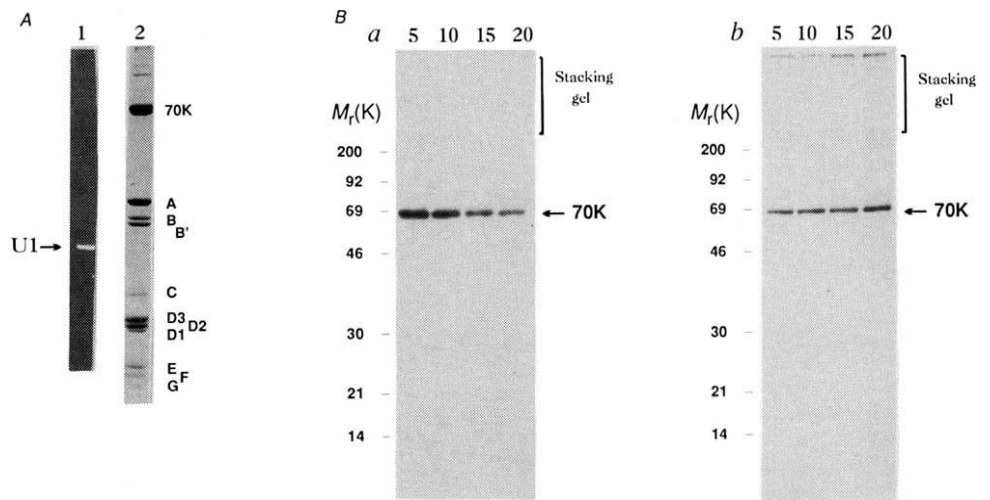
Figure 1A shows the protein composition of purified U1 snRNP. Highly purified U1 snRNPs from human cells contain a protein kinase activity that phosphorylates the 70K protein selectively *in vitro*. Phosphopeptide analysis reveals that the 70K protein phosphorylated *in vitro* has several phosphopeptides in common with the protein phosphorylated *in vivo*, indicating that the kinase associated with isolated U1 snRNPs may be the same as the one acting *in vivo* (U.K. *et al.*, manuscript in preparation).

To investigate the effect of 70K phosphorylation on the function of U1 snRNP in the splicing reaction, we phosphorylated the 70K protein of isolated U1 snRNPs using ATP- γ S as a thiophosphate donor. We then used these thiophosphorylated U1 snRNPs for complementation of nuclear splicing extracts

Received 11 January; accepted 10 March 1993.

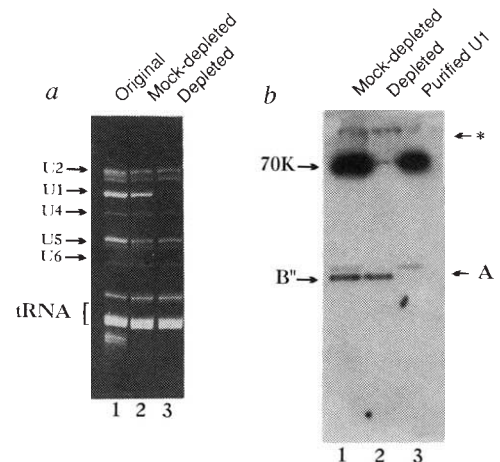
1. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. *Science* **253**, 49-53 (1991).
2. Finlay, C. A., Hinds, P. W. & Levine, A. J. *Cell* **57**, 1083-1093 (1989).
3. Zambetti, G. P., Bargonetti, J., Walker, K., Prives, C. & Levine, A. J. *Genes Dev.* **6**, 1143-1152 (1992).
4. Farmer, G. *et al. Nature* **358**, 83-86 (1992).
5. Baker, S. J., Markowitz, K., Fearon, E. R., Willson, J. K. V. & Vogelstein, B. *Science* **249**, 912-915 (1990).
6. Ginsberg, D., Mechta, F., Yaniv, M. & Oren, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9979-9983 (1991).
7. Lechner, M. S. *et al. EMBO J.* **11**, 3045-3052 (1992).
8. Chin, K.-V., Ueda, K., Pastan, I. & Gottesman, M. M. *Science* **255**, 459-462 (1992).
9. O'Shea-Greenfield, A. & Smale, S. T. *J. Biol. Chem.* **267**, 1391-1402 (1992).
10. Ptashne, M. *Science* **335**, 683-698 (1988).

FIG. 1 Protein composition of purified U1 snRNP (A) and dephosphorylation in HeLa nuclear extracts of a phosphorylated or thiophosphorylated 70K protein (B). A, U1 snRNP was purified as described²³. Lane 1 shows the fractionation in a denaturing gel of the RNAs contained in purified U1 snRNP; lane 2 shows the analysis by SDS-polyacrylamide gel electrophoresis of the proteins contained in purified U1 snRNP. RNAs and proteins were stained by ethidium bromide (lane 1) and Coomassie blue (lane 2), respectively. B, Purified U1 snRNPs (10 μ g) were incubated for 30 min in the presence of 100 μ Ci of either ³²P-labelled ATP (a) or a ³⁵S-labelled ATP- γ S (b) in buffer D²⁴ in a total volume of 60 μ l. To remove unreacted radioactive nucleotides, reaction mixtures were passed over a 1 ml Sephadex-G50 column. Aliquots (20 μ l) from the peak fractions containing radiolabelled U1 snRNPs were added to a reaction mixture containing in a final volume of 100 μ l 2 mM ATP, 20 mM creatine phosphate, 3.2 mM MgCl₂, 2.6% polyvinylalcohol and 40 μ l nuclear extract prepared as described²⁴. After incubation at 30 °C for the time intervals



shown, 10- μ l samples were withdrawn and applied to a 12% SDS-polyacrylamide gel. Autoradiograms of the fractionated proteins are shown in a and b. Size markers are indicated on the left of each panel and correspond to the position of pre-stained proteins (Amersham Rainbow markers).

FIG. 2 Removal of endogenous U1 snRNP from splicing extract. a, An RNA gel, showing the various RNAs contained in the untreated splicing extract (lane 1), and splicing extracts treated with streptavidin-agarose in the absence (lane 2) or presence (lane 3) of 5' splice-site RNA. Densitometry of the northern blot autoradiograph reveals that the amount of U1 snRNA in the depleted extract represents less than 1% of that in mock-depleted extract (data not shown). b, Immunoblot of proteins in splicing extracts treated with streptavidin-agarose in the absence (lane 1) or presence (lane 2) of 5' splice-site RNA. The blot was initially probed with a monoclonal antibody (antibody 2.73) against U1 70K as described²⁵ and then subsequently with a monoclonal antibody (4G3) against U2 B'' protein²⁶. As a control, proteins from purified U1 snRNPs were used as markers (lane 3). Cross-reacting proteins are indicated on the right; these were stained with antibody 2.73 (asterisk) or 4G3 (U1-specific A protein). METHODS. The m⁷GpppG-capped 5' splice-site RNA¹¹ was biotinylated²⁷ and coupled to presaturated streptavidin-agarose²⁸. Unbound RNA was removed by extensive washing of agarose beads in Net 1/2 buffer (25 mM Tris-HCl, pH 7.5, 50 mM NaCl, 4 mM MgCl₂, and 0.025% NP40). For the U1 snRNP depletion, 250 μ l HeLa cell nuclear extract in buffer 1 (20 mM triethanolamine, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 0.8 mM ATP, 16.6 mM creatine phosphate, 2.6 mM MgCl₂, 0.5 mM DTT and 20% glycerol), which had been incubated for 30 min at 30 °C, was mixed with 250 μ l streptavidin-agarose containing 50 μ g 5' splice-site RNA in buffer 1 and rotated end-over-end for 2 h at 4 °C. The U1 snRNP-depleted extract was separated from the agarose beads by centrifugation. As a control, depletion was carried out in the absence of the 5' splice-site RNA.



that had been depleted of endogenous U1 snRNPs.

Figure 1B shows that the kinase activity associated with U1 snRNP accepts ATP- γ S as a substrate and thiophosphorylates the U1 70K protein efficiently; also, the thiophosphate groups remain stably attached to the 70K protein and are resistant to removal by phosphatases in HeLa cell splicing extracts (Fig. 1B, b). In contrast, normal phosphate groups attached to U1 snRNP 70K protein are slowly removed (Fig. 1B, a).

For complementation studies it was essential to use splicing extracts from which the endogenous U1 snRNPs and free 70K protein had been removed completely. An artificial RNA containing the consensus 5' splice-site sequence forms a stable complex with U1 snRNP under low-salt conditions¹³, suggesting that this RNA could be used to deplete nuclear extracts of U1 snRNPs. The m⁷G-capped RNA was biotinylated chemically at both ends, immobilized on streptavidin-agarose and incubated with splicing extracts under low-salt conditions at 4 °C. Removal of U1 snRNPs from nuclear extracts was more than 98% com-

plete, as judged by RNA gel electrophoresis (Fig. 2a; compare lanes 2, 3). Furthermore, immunoblot analysis of the U1-depleted nuclear extracts revealed that endogenous 70K protein as well as the U1 snRNP-specific protein A were quantitatively removed (Fig. 2b; compare lanes 1, 2). As a control, U2 snRNP-specific B'' protein was not affected by our depletion procedure (Fig. 2b; compare lanes 1, 2).

As expected, the nuclear extract depleted of U1 snRNP had no splicing activity (lanes 8 in Fig. 3a and b). Splicing could be fully restored, however, upon addition of purified U1 snRNPs (Fig. 3a, lane 7), demonstrating that splicing factors other than U1 snRNP were present in the U1-depleted extract and had not been removed by our depletion procedure.

Complementation of U1-depleted nuclear extracts with purified U1 snRNPs that had been phosphorylated with normal ATP also fully restored the splicing reaction (Fig. 3a, lane 6). In contrast, the U1 snRNPs that had been thiophosphorylated with ATP- γ S completely failed to complement the splicing

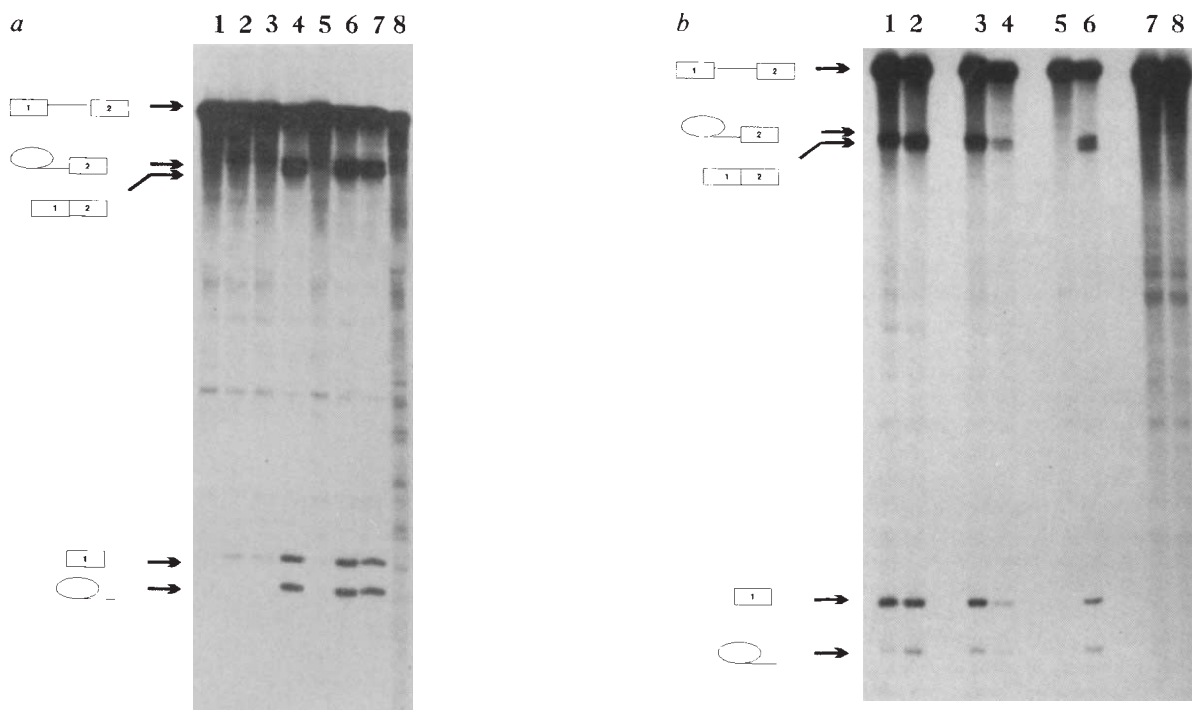


FIG. 3 *a*, U1 snRNPs containing thiophosphorylated U1 70K are unable to complement depleted extracts. Splicing complementation was done in 12- μ l reactions consisting of 32 P-labelled SP6 pre-mRNA containing the first two exons and first intron of wild-type human β -globin, 2 mM ATP, 20 mM creatine phosphate, 3.2 mM MgCl_2 , 2.6% polyvinyl alcohol, 6 μ l U1 snRNP-depleted nuclear extract (lanes 5–8), which were preincubated for 10 min at 30 °C in the presence of either 0.5 mM ATP (lane 4) or 0.5 mM ATP- γ S (lanes 1–3), and 3 μ l untreated (lanes 3, 4, 7), phosphorylated (lanes 2, 6) or thiophosphorylated (lanes 1, 5) purified U1 snRNPs (corresponding to 100 ng U1 snRNA). In the control, U1 snRNP was replaced with 3 μ l buffer D (lane 8). After incubation for 60 min at 30 °C, pre-mRNA splicing products were extracted from 7 μ l of each reaction and separated on a 6% polyacrylamide-urea gel. Phosphorylation of the U1-specific 70K protein was achieved using either 2 mM ATP (lanes 2, 6) or 2 mM ATP- γ S (lanes 1, 5) following the same procedure as for Fig. 1*B*, except that no labelled ATPs were added.

After removal of unreacted nucleotides on Sephadex-G50, we confirmed that serine phosphorylation of the 70K protein was nearly quantitative, as further incubation of the U1 snRNPs with 32 P-labelled ATP gave no significant incorporation of 32 P label. *b*, Inhibition of splicing caused by thiophosphorylation of U1 snRNP can be relieved by an excess of phosphorylated U1 snRNP. To accommodate additional U1 snRNPs, splicing reactions were done in 20 μ l under the splicing conditions already described but with the following additions of phosphorylated and thiophosphorylated U1 snRNPs (corresponding to 30 ng μ l $^{-1}$ U1 snRNA): lane 1, 8 μ l phosphorylated U1 snRNP; lane 2, 6 μ l phosphorylated and 2 μ l thiophosphorylated U1 snRNPs; lane 3, 4 μ l phosphorylated U1 snRNP; lane 4, 2 μ l phosphorylated and 2 μ l thiophosphorylated U1 snRNPs; lane 5, 2 μ l thiophosphorylated U1 snRNP; lane 6, 2 μ l phosphorylated U1 snRNP. Lanes 7 and 8 are control experiments without U1 snRNP, done either in the presence of ATP- γ S or of ATP, respectively.

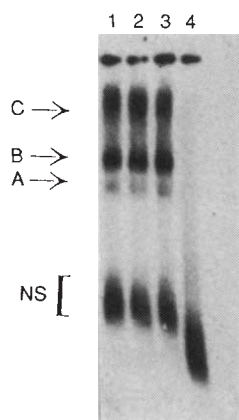


FIG. 4 Spliceosome formation with thiophosphorylated U1 snRNPs. Aliquots (5 μ l) from the same splicing reactions as those corresponding to lanes 5–8 in Fig. 3 (lanes 1–4 respectively) were withdrawn after 30 min of incubation, treated with heparin (2 mg ml $^{-1}$), mixed with 1 μ l of 97% glycerol containing 1% bromophenol blue and then resolved on a non-denaturing agarose-polyacrylamide composite gel²⁹. The time course of spliceosome assembly was not affected by the state of phosphorylation of the 70K protein (data not shown).

reaction (Fig. 3*a*, lane 5). Both steps of splicing were inhibited in the presence of thiophosphorylated U1 snRNP. Inhibition of splicing by a contaminant in the thiophosphorylated U1 snRNP preparation can be ruled out, because adding an excess of phosphorylated U1 snRNP restores splicing activity (Fig. 3*b*, lanes 2 and 4).

We have previously shown that replacement of ATP in normal splicing extracts by ATP- γ S leads to the inhibition of the second step of splicing, implying that the state of phosphorylation of a splicing factor other than U1 70K is also critical for the splicing reaction¹⁴. To investigate the inhibitory effect of 70K protein thiophosphorylation, we incubated U1 RNP-depleted extracts with ATP- γ S to allow thiophosphorylation of splicing factors other than U1 snRNP by the endogenous kinase. In agreement with our previous findings, when either phosphorylated or non-phosphorylated U1 snRNPs were added to this extract, the first step of splicing took place, whereas the second was completely inhibited (Fig. 3*a*, lanes 2, 3). But when thiophosphorylated U1 snRNPs were added to the pre-treated extract, neither step of the splicing reaction occurred (Fig. 3*a*, lane 1). These results support the idea that the inhibition of splicing upon thiophosphorylation of the U1 70K protein is directly related to the function of the 70K protein in the first step of splicing. They also show that the function of at least one other splicing factor is sensitive to thiophosphorylation¹⁴.

We have used native gel electrophoresis to show that formation of mature spliceosomes is not inhibited in the presence of thiophosphorylated U1 snRNPs in splicing extracts (Fig. 4). It

is therefore most likely that the thiophosphorylation of the 70K protein affects a pre-catalytic step of the splicing reaction.

Our data show that the 70K protein is important for the splicing reaction *in vitro* and, moreover, that its function depends critically upon the way in which it is phosphorylated. But the molecular basis of the inhibition of splicing by thiophosphorylation of the 70K protein is unclear. It may be that the 70K protein has to undergo a phosphorylation-dephosphorylation cycle during its action in the spliceosome. This cycle would be blocked by the failure of phosphatases to release the phosphorothioate thiophosphorylated groups from the 70K protein. This idea seems attractive and is supported by the finding that highly specific inhibitors (okadaic acid, tautomycin and microcystin-LR) of protein phosphatases PrP1 and PrP2a differentially inhibit the two catalytic steps of mammalian pre-mRNA splicing^{14,15}. Inhibition of PrP2a predominantly inhibits the second catalytic step of splicing¹⁴, whereas inhibition of PrP1 and PrP2a affects both splicing steps¹⁵. A form of PrP1 could be a good candidate for the phosphatase involved in U1 70K dephosphorylation. In addition, and this may not be mutually exclusive, the phosphate groups could function as part of a recognition site on the 70K protein for other spliceosomal factors; this recognition would be disturbed if the normal phosphate were replaced by a phosphorothioate group.

The kinase activity associated with the U1 snRNP selectively phosphorylates the arginine/serine rich domains (R/S domains) in the C-terminal half of the 70K protein (our unpublished results). Interestingly, R/S domains have also been found in many metazoan non-snRNP splicing factors¹⁶⁻²². Preliminary evidence indicates that other splicing factors containing R/S domain (such as SF-2 and U2AF) are also phosphorylated^{16,19,20}. Investigation of the importance of phosphorylation in the function of the U1 70K protein should eventually provide us with information on the role of the R/S domains of other splicing factors. Furthermore, as U1 snRNPs and the aforementioned splicing factors are involved in the regulation of alternative splicing¹⁷⁻²⁰, their post-translational phosphorylation may finely tune the splicing reaction. □

Received 3 September 1992; accepted 17 February 1993.

1. Steltz, J. A., Black, D. A., Gerke, V. & Parker, K. A. in *Structure and Function of Major and Minor snRNPs* 115-154 (Springer, Berlin, 1988).
2. Lührmann, R., Kastner, B. & Bach, M. *Biochim. biophys. Acta* **1087**, 265-292 (1990).
3. Green, M. R. *A. Rev. Cell Biol.* **7**, 559-599 (1991).
4. Ruby, S. W. & Abelson, J. *Trends Genet.* **7**, 79-85 (1991).
5. Guthrie, C. *Science* **253**, 157-163 (1991).
6. Rymond, B. C. & Rosbash, M. in *The Molecular and Cellular Biology of the Yeast Saccharomyces* (Cold Spring Harbor Laboratory Press, New York, 1992).
7. Michaud, S. & Reed, R. *Genes Dev.* **5**, 2534-2546 (1991).
8. Jamison, S. F. & Garcia-Blanco, M. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 5482-5486 (1992).
9. Tatei, K., Takemura, K., Tanaka, H., Masaki, T. & Hoshima, V. *J. biol. Chem.* **267**, 11667-11674 (1992).
10. Heinrichs, V., Bach, M., Winkelmann, G. & Lührmann, R. *Science* **247**, 69-72 (1990).
11. Wooley, J. C., Zuckenberg, L. R. & Chung, S. Y. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5208-5212 (1983).
12. Woppmann, A., Patschinsky, T., Bringmann, P., Godt, F. & Lührmann, R. *Nucleic Acids Res.* **18**, 4427-4438 (1990).
13. Khellil, S., Daugeron, M. C., Alibert, C., Jeanteur, P., Cathala, G. & Brunel, C. *Nucleic Acids Res.* **19**, 877-884 (1991).
14. Tazi, J., Daugeron, M. C., Cathala, G., Brunel, C. & Jeanteur, P. *J. biol. Chem.* **267**, 4322-4326 (1992).
15. Mermoud, J. E., Cohen, P. & Lamond, A. I. *Nucleic Acids Res.* **20**, 5263-5269 (1992).
16. Zamore, P. D., Patton, J. G. & Green, M. R. *Nature* **355**, 609-614 (1992).
17. Amrein, H., Gorman, M. & Nothiger, R. *Cell* **55**, 1025-1035 (1988).
18. Gorski, T. J., Edstrom, J. E. & Baker, B. S. *Cell* **56**, 1011-1018 (1989).
19. Ge, H., Zuo, P. & Manley, J. L. *Cell* **66**, 373-382 (1991).
20. Krainer, A. R., Mayeda, A., Kozak, D. & Binns, G. *Cell* **66**, 383-394 (1991).
21. Fu, X. D. & Maniatis, T. *Science* **256**, 535-538 (1992).
22. Zahler, A. M., Lane, W. S., Stolk, J. A. & Roth, M. B. *Genes Dev.* **6**, 837-847 (1992).
23. Bach, M., Bringmann, P. & Lührmann, R. *Meth. Enzym.* **181**, 232-256 (1990).
24. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. *Nucleic Acids Res.* **11**, 1475-1489 (1983).
25. Lelay-Taha, M. N., Révellaud, I., Sri-Wridada, J., Brunel, C. & Jeanteur, P. *J. molec. Biol.* **189**, 519-539 (1986).
26. Habets, W. J., Hoet, M. H., De Jong, B. A. W., van der Kemp, A. & van Venrooi, W. J. *J. Immun.* **143**, 2560-2566 (1989).
27. Tazi, J., Forné, T., Jeanteur, P., Cathala, G. & Brunel, C. *Molec. cell. Biol.* (in the press).
28. Blencowe, B. J., Sproat, B. S., Ryder, U., Barabino, S. & Lamaond, A. I. *Cell* **59**, 531-539 (1989).
29. Nelson, K. K. & Green, M. R. *Genes Dev.* **2**, 319-329 (1988).

ACKNOWLEDGEMENTS. We thank I. Öchsner-Welpelo, S. Brner and R. Dietz for technical assistance. This work was supported by grants from the Deutsche Forschungsgemeinschaft and the Fonds der Chemischen Industrie (R.L.) and from the Centre National de la Recherche Scientifique, the Association pour la Recherche contre le Cancer and the University of Montpellier II (C.B.).

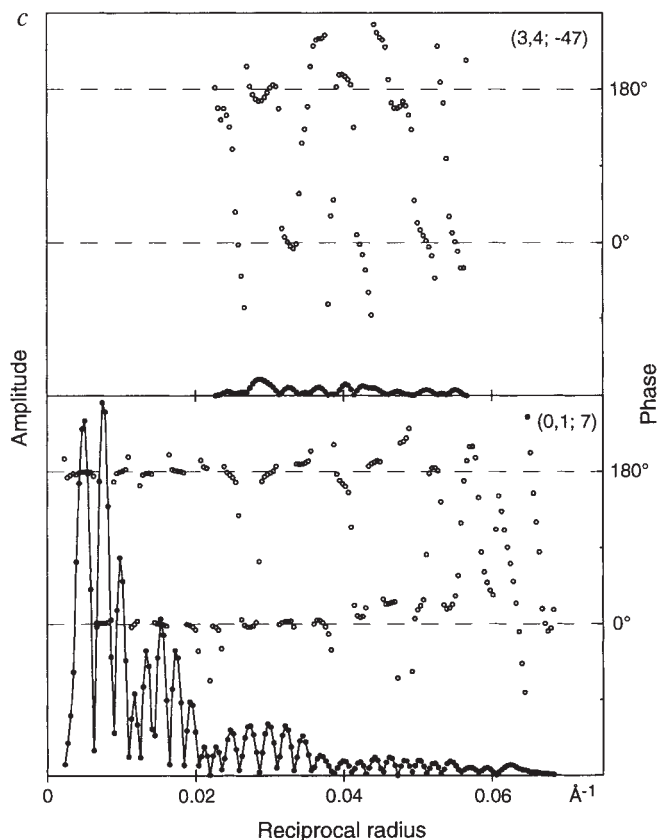
ERRATA

Three-dimensional cryo-electron microscopy of the calcium ion pump in the sarcoplasmic reticulum membrane

Chikashi Toyoshima, Hiroyuki Sasabe & David L. Stokes

Nature **362**, 469-471 (1993)

FIGURE 1c was omitted from the above letter during the production process and is now shown below.



In addition, part of the label at the bottom left-hand corner of Fig. 4a was unclear. This should read 'A2(M6, M8)'. □

Dispersion of neural progenitors within the germinal zones of the forebrain

Gord Fishell, Carol A. Mason & Mary E. Hatten

Nature **362**, 636-642 (1993)

THE date of receipt of this letter was erroneously given as 14 December 1992 instead of 2 September 1992. □